2013

Unpacking the Study of Instructional Improvement: Issues, Outcomes, and Implications of Three Comprehensive School Reform Efforts

Gabriella Silva Gorsky
Loyola University Chicago, g.gorsky@yahoo.com

LOYOLA UNIVERSITY CHICAGO

UNPACKING THE STUDY OF INSTRUCTIONAL IMPROVEMENT: ISSUES, OUTCOMES, AND IMPLICATIONS OF THREE COMPREHENSIVE SCHOOL REFORM EFFORTS

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF ARTS

PROGRAM IN RESEARCH METHODOLOGY

BY

GABRIELLA S. GORSKY

CHICAGO, IL

AUGUST 2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

There is a long history of educational reform efforts in the United States. Current literature on school reform suggests that a comprehensive approach that involves implementing instructional change across many instructional domains is more effective at producing the types of far-reaching improvement desired compared with mono-focal approaches focusing on a single new strategy. This study explored the impacts and outcomes of the Accelerated Schools Project (ASP), America's Choice (AC), and Success for All (SFA) within the Study of Instructional Improvement (Ball, Cohen, & Rowan, 2010) schools to answer three primary questions: Are student outcomes on *TerraNova* mirrored by outcomes on *Supera* (the Spanish language *TerraNova* test)? How do the three CSR programs compare in terms of how they impact student growth over time based on *TerraNova* test scores? What are the effects of student- and school-level factors on the outcomes of these comprehensive school reform initiatives? To answer these questions the researcher first conducted a series of *t*-tests and then built a multi-level growth model. The analyses demonstrated that while it is clear that differences exist between students, the true impact of the CSR programs is more ambiguous. It may be the case that the student- and school-level characteristics, present from the outset, predict academic growth more than actual program participation.

CHAPTER ONE

INTRODUCTION

For over a century, K-12 school systems across the United States have implemented a plethora of Comprehensive School Reform (CSR) initiatives in an effort to produce meaningful academic growth in their students. The last two decades are certainly no exception; in the era of the No Child Left Behind Act (NCLB), stakes are high, and teachers, administrators, school board members, and other advocates for student success are as eager as ever to put effective improvement programs into practice. Some of these programs target literacy and some target Mathematics skills (Mumme & Weissglass, 1990; Tivman & Hemphill, 2005). Regardless of their articulated foci, many aim to better prepare students for standardized testing. This, in particular, is an area of much debate and one that is shrouded in controversy. Test scores are typically the measuring stick against which student success, and especially student improvement, are evaluated and there is an immense amount of pressure to produce meaningful yearly gains (in some cases even biannually or quarterly). When a school has adopted any given CSR intervention the expectation for success only increases, as does the need for a valid and unbiased evaluation of scholastic outcomes.

According to Ball, Cohen, and Rowan (2010), the Comprehensive School Reform (CSR) interventions borne of NCLB were divergent from earlier initiatives, which tended to focus on making changes to a single aspect of the educational approach. These earlier narrow-sighted programs, which might simply introduce a new textbook into the curriculum or provide additional academic support services to “at-risk” students, rarely result in any kind of meaningful widespread change. New reform efforts approached instructional improvement on a broader scale, creating changes all across the educational landscape of the school. From training teachers in new curricular strategies, to reaching out to parents for additional support and guidance, to integrating program staff into the school environment, by addressing the many intersecting components of teaching and instruction, the three CSR initiatives investigated in the SII aimed to multiply instructional scope at the school level.

While looking at the impact of a program in a single school might provide some meaningful information for that sample, it does not necessarily paint a complete picture when it comes to how effective the program will be in another academic setting. One way to evaluate the efficacy of CSR initiatives on a broader scale is to compare the student outcomes from a variety of schools that have implemented the same specific CSR intervention. From 2000 to 2004, researchers from the University of Michigan at Ann Arbor designed and implemented a longitudinal quasi-experiment called the Study of Instructional Improvement (SII) for which they collected student and teacher data from 89 schools that had undertaken one of the three most popular CSR interventions for elementary and middle school settings in the United States: Accelerated Schools Project (ASP), America’s Choice (AC), or Success for All (SFA) (Ball et al., 2010).

Each of these programs has a unique theoretical orientation, as well as distinct strategies for administering school-wide change. Students’ academic outcomes were measured using *TerraNova* test scores collected in the fall and the spring for two cohort groups during the same academic years: Cohort A (Kindergarten to 2nd grade) and Cohort B (3rd to 5th grade). By looking at the relationship between student- and school-level variables and the academic outcomes of students in each program, it becomes possible to begin to untangle the ways each program does or does not work. Of particular interest is who the programs are working for and who they may be neglecting

CHAPTER TWO

LITERATURE REVIEW

Over a decade after No Child Left Behind (NCLB) was passed, schools across the country are still struggling to meet rigorous state achievement standards. According to Pianta, Belsky, Vandergrift, Houts, and Morrison (2008), the quality of early instructional experiences can have a long-lasting impact on students' achievement trajectories. They explored the way the quality of instructions impacts student achievement and found that many components of the educational environment from curricular materials to the teacher-student relationship can have a noteworthy effect on academic growth. This suggests that it is imperative that effective comprehensive initiatives are available for schools in need of reform.

No Child Left Behind (NCLB) and the Achievement Gap

There is no question that the passing of the No Child Left Behind Act in 2001 had a tremendous impact on the K-12 educational landscape. This was especially true for minority students who made up the greatest percentage of those considered "at-risk" and/or attending struggling schools. Many researchers have explored the role of NCLB in the lives of minority students and the impact it has had on their achievement. In particular, there is a focus on articulating how NCLB has affected the achievement gap that exists between minority and non-minority students (Gorey, 2009; Kyung, 2011; Reardon, Greenberg, Kalogrides, Shores, & Valentino, 2012; Slavin & Madden, 2001).

Despite having a purported aim of improving the academic performance of all students through increased access to “quality” educational services, NCLB has still left many students behind, especially those belonging to minority and/or economically disadvantaged populations.

There is evidence that NCLB has had a positive impact on student achievement, but results are mixed. A report by the U.S. Department of Education revealed that students attending Title I schools (those most in need of improvement as a result of failing to meet standards for two or more years) who enrolled in supplementary educational services did make progress, especially in reading and Mathematics (Zimmer, Gill, Razquin, Booker, & Lockwood, 2007). However, this investigation also revealed that students who chose to leave their Title I school and attend another school rather than utilize educational support services did not see the same achievement gains, which suggests that not all outcomes of NCLB have been positive and raises questions about the factors contributing to student success. Though more states are meeting standards and making claims that the achievement gap is closing (Blank, 2011), the reasons for this are unclear and allegations of lowered standards, teaching to the test, and even cheating (from providing students with test answers to changing students answer sheets) are more prevalent than ever (Burke, 2012). Clearly, all students, but especially racial and ethnic minority and economically disadvantaged students are at risk under the weight of NCLB mandates and seemingly near-unattainable expectations.

Opponents of the high-stakes testing movement often base their opposition on the phenomenon of “teaching to the test,” where instruction time is transformed into test-prep time, often many months before the exam is even scheduled to take place. In his article on the impact of high-stakes testing on curriculum, Au (2007) describes the way NCLB shifted the emphasis away from meaningful learning experiences and towards a test-driven pedagogical approach. Based on a qualitative meta-synthesis of studies examining the impact of standardized testing on curricular practices and outcomes, Au explains that the overall trend is towards a narrowing of content to conform to test subjects. Beyond that, he noted that content is often taught as dissected pieces as opposed to broad ideas, and that instruction becomes teacher-centered, as educators struggle to cover the breadth of material needed to prepare students for exams. However, Au also found that in a small number of cases, high-stakes testing could lead to a more student-centered learning environment where content is not narrow, but expanded, and content is not fragmented, but cohesive. Though this suggests that NCLB and high-stakes testing do not necessarily inhibit schools' ability to exert curricular control, the question of whether or not a test-driven curriculum is a bad thing for students, teachers, or schools remains.

Comprehensive School Reform is the Key

Overwhelmingly, the solution to the problems plaguing Title I schools, as well as those merely looking to improve overall achievement and bring more students up to meet state and federal standards, lies in Comprehensive School Reform (CSR). With higher performance expectations, CSR programs provided a foundation for creating lasting change within schools. NCLB mandates calling for a comprehensive redesign of academic approaches, including instructional practices, the choice of assessment tools, and the role of professional development for teachers and administrators has led to the implementation of a variety of CSR approaches, especially in the years directly following NCLB (Tushnet, Flaherty, & Smith, 2004).

Across the country, schools and school districts incorporating CSR initiatives into their plans for encouraging far-reaching improvements in student achievement have seen enormous gains, especially when it comes to mathematics and reading. Mac Iver and Mac Iver (2009) found that Philadelphia schools employing a CSR program that incorporated a mathematics curriculum saw dramatic gains in mathematics scores on the Pennsylvania System of School Assessment and that this growth was directly related to the number of years each school had incorporated the mathematics content into the revised curriculum. Though this was not always the case (Evans-Andris & Usui, 2008), most studies indicated that CSR interventions have a significant positive impact on student achievement and can provide useful organizational strategies for instructional improvement (Rowan & Miller, 2007). However, the question of which formal CSR program is most impactful still remains unanswered.

Accelerated Schools Project (ASP)

The Accelerated Schools Project (ASP) (Levin, 1986) is based on the concept of “powerful learning” and it is this abstract concept that guides the intervention. The central focus of the program is on cultivating differentiated and student-centered instruction that will lead to “powerful learning” for the students. According to Ball et al. (2010), for the duration of SII, ASP used an approach with broad and general program goals. Rather than explicitly directing teachers in what sorts of changes to make, each school was expected to develop their own set of strategies for improvement based on the specific school context and the unique challenges faced by the student population.

Byrd and Finnan (2003) explain that the ASP takes a teach-to-the-top approach in which students, especially those at risk, are taught an accelerated curriculum. A core goal is to bring the lowest achieving students up to meet grade-level standards before transitioning into high school. ASP was designed to enrich the learning experiences of all students and improve performance on state standardized tests through school-wide collaboration. According to Byrd and Finnan, schools utilizing the ASP approach not only saw score gains across all subjects, but also saw a reduction in disciplinary actions and an improvement in student attendance. Despite the improvement described in Byrd and Finnan’s study, little research exists on the impact of ASP so it is unclear if these results are truly representative of intervention outcomes.

Success for All (SFA)

Success for All (SFA) (Slavin, 1987-88) was designed around the concept of instructional “routines” for teaching reading and provides discrete and specific guidelines for instructional improvement. According to the Success for All website, the program currently serves over 1,000 schools in forty-seven states (Success for All, 2012). Ball et al. (2010) explain that SFA schools involved in the SII were required to appoint a literacy coordinator and several other liaison positions to facilitate the implementation of the intervention. As a result, SFA schools were very high in instructional leadership and had school-wide reform plans that were unambiguous and diligently applied in each school setting.

According to a meta-evaluation of the Success for All reform model by the What Works Clearinghouse (2009) in conjunction with the U.S. Department of Education, SFA was found to have positive effects on students' reading achievement, especially when it came to alphabetics and comprehension. However, findings of this investigation also reveal that these specific content areas were more strongly impacted than others (i.e., reading fluency). Second year outcomes for the national field trial of SFA revealed significant school-level effects for four reading domains: Letter Identification, Word Identification, Word Attack, and Reading Comprehension. While all were found to be improved under SFA implementation, Word Attack was impacted least and had the greatest within-school variation (Borman et al., 2005).

According to an internal report comparing Success for All to another initiative (Roots & Wings, 1996), gains can be seen across all academic areas, but this depends heavily on how broadly and appropriately the program is implemented in a school setting (Slavin & Madden, 2004). Of particular interest is their finding that SFA produced marked improvement in English-Language Learners (ELL), in particular, Spanish-speaking students in Arizona. However, the authors also note that improvement was not statistically significant for all measures in every study. Additionally, studies have revealed that low socio-economic status (SES) and low-achieving students receive the most benefit from the SFA program, leading to reduced special education placement (Slavin & Madden, 2004; Hurley, Chamberlain, Slavin, & Madden, 2000).

America's Choice (AC)

The America's Choice (AC) (NCEE, 1988) program was designed as a response to the standards-based school reform movement by the National Center on Education and the Economy (NCEE). According to NCEE, this intervention was being used by over 500 schools by 2004. As a result, unlike ASP, which was more abstract in its implementation, AC delineates clear goals and strategies for improving student performance. In particular, the program focuses on developing students' writing skills first before moving on to reading and mathematics. Ball et al. (2010) explain that for the SII, teachers in schools utilizing the AC model were not only provided with a curriculum guide, but also participated in training programs aimed at developing specific writing routines for use in the classroom. In addition, AC coaches and facilitators played an active role in assisting teachers in the formulation of rubrics and other guidelines for assessing written work.

In an external evaluation of the AC program model, Poglinco et al. (2003) explored the role of coaches in the implementation of the specific strategies and tools designed to increase achievement. While AC does address mathematics content, the focus is on the development of language arts skills (writing and reading) with the ultimate goal of bringing all students up to meet clearly articulated standards. According to Poglinco et al., AC coaches play a crucial role in the implementation of the program, closely monitoring teacher practices and ensuring that the learning strategies are properly conducted in the classroom. As a result, AC schools tended to be very high in instructional leadership (Ball et al., 2010).

Whereas ASP targets “at-risk” students, the AC approach is grounded in high expectations for all students and requires school-wide consensus on participation. In order to become an AC school, 80% of faculty must commit to a three-year implementation of the study (Supovitz, Taylor, & May, 2002). In addition to coaches, a literacy coordinator and parent/community liaison are required in order to ensure that students and teachers in AC schools are receiving the support required to produce the desired student achievement gains (Ball, Cohen, & Rowan, 2010; Supovitz et al., 2002). In an evaluation of middle schools implementing the AC initiative, Supovitz et al. (2002) found that students in AC schools outperform students in other schools on writing. However, they also noted that these students’ performance in reading and mathematics was less distinct from control schools. This outcome is no surprise given the emphasis AC places on developing students’ writing skills and suggests that SII schools utilizing the AC model may not have experienced as much growth in other subject areas as those utilizing ASP or SFA. This may have been especially true for reading since it is the core focus of SFA.

The Importance of Language Equivalence

Schools in communities with high minority populations or those serving impoverished areas may have a need for Spanish-language versions of standardized test instruments. While these tests are carefully developed for equivalence with the English language version, this may not always be reflected in student performance. In the present study it is understood that ASP, AC, and SFA all focus on improving the performance of academically “at-risk” students. In many cases, students requiring the use of *Supera*

(Spanish language *TerraNova* test) represent those with the greatest need. They may be English Language Learners (ELL) struggling to follow the lessons in class, recent immigrants whose prior educational experiences are not equivalent to grade-level expectations at the new school, or those with parents who do not speak English and who may not have attained a very high level of education.

If student performance is going to be measured using a translated assessment tool, it is imperative that there is equivalence between that tool and the original version. According to van de Vijver and Tanzer (1997), without equivalence of measures, defined as a lack of bias, valid comparisons cannot be made across cultural populations. Given the importance of academic testing outcomes, it is imperative that students' abilities are equitably assessed. In a study evaluating the psychometric properties of a Spanish-language version of the Social Skills Scale, Jurado, Cumba-Aviles, Colazzo, and Matos (2012) reinforced the idea that assessment instruments must be viewed as the product of a specific cultural context, and reflect the norms, values, and attitudes of that culture.

Unlike construct and method bias, which occur at various levels of test administration, item bias occurs when the content of a particular question carries different meaning across cultures (van de Vijver &Tanzer, 1997). When this happens, certain individuals are at an advantage in answering while others are at a disadvantage, and the result is that the test is no longer capturing what it claims; rather, it is measuring the test-taker's familiarity with the cultural context within which the tool was developed. For non-English speaking students, the negative impact of item bias on standardized test performance can have long-lasting effects on their academic and personal lives and should be considered in the evaluation of testing materials.

Purpose

In light of the current high-intensity educational landscape, it is imperative that a school can count on the success of a CSR program. An intervention that turns out to be a bad fit or doesn't produce notable improvements can mean lost time and money in an epoch where school districts nationwide are short on both. When deciding on an initiative to utilize, it makes sense that those programs which are most popular and accepted within the academic community would be likely choices. School administrators could review dozens of reports of the success of ASP, AC, or SFA. However, these would merely be evaluations of a *single* program within a school or district, not a comparison of multiple interventions to one another.

While it is valuable to know the impact of an individual program within one school or system, a more complete picture of how different initiatives unfold within a variety of school settings can provide school districts and administrators with a stronger basis for making an educated decision regarding which program will produce tangible change within their school(s). By including factors such as school type (i.e., charter, magnet, or public), the SES of the local community of each school (the Community Disadvantage Index), and the racial/ethnic breakdown for each program, it becomes possible to match "at-risk" schools to the program(s) that will work best for them.

Currently, little literature exists that does a cross-intervention contrast of student achievement following the implementation of a CSR program. The purpose of the investigation is first to describe the performance of each of the three programs evaluated in the Study of Instructional Improvement (Ball, Cohen, & Rowan, 2010) using test score data and then to tell the story of the success and failures of each program during the period of data collection. The goal is to arrive at an understanding of the factors impacting the effectiveness of each program so that ASP, AC, and SFA can be reasonably compared by those interested in adopting a CSR program. This study aims to answer three fundamental questions to establish parallelism between the two test versions, differences in achievement gains across CSR programs, and differences between individuals and schools that mediated the potential for success: Are student outcomes on *TerraNova* mirrored by outcomes on *Supera* (the Spanish language *TerraNova* test)? How do the three CSR programs (ASP, AC, and SFA) compare in terms of how they impact student growth over time based on *TerraNova* test scores? What are the effects of student- and school-level factors on the outcomes of these comprehensive school reform initiatives?

When it comes to how *TerraNova* scores will align with *Supera* scores, research on the impact of item bias and test non-equivalence suggests English language test-takers will outperform Spanish language test-takers in both mathematics and reading, even while holding other variables constant due to the role specific cultural context plays in the formulation of test items and the impact this has on students' ability to answer correctly (Jurado et al., 2012; van de Vijver &Tanzer, 1997).

Drawing on results from evaluations of the three reform initiatives, it is expected that overall gains will be greater for reading scores than for mathematics scores (Supovitz et al., 2002; What Works Clearinghouse, 2009). Specifically, it is expected that SFA will lead to the greatest reading gains among the three reforms because of the emphasis the program places on improving reading skills (Borman et al, 2005, What Works Clearinghouse, 2009). Though the other two CSR initiatives have reading components, AC places a greater emphasis on developing writing skills specifically (Supovitz et al., 2002), while the broad and general program guidelines for ASP means that different schools utilizing the program may have prioritized different subject areas (Ball et al., 2010). Though mathematics achievement is generally less emphasized by the CSR programs than literacy achievement, evidence from Byrd and Finnan (2003) suggested that ASP produces score gains across all subject areas. While Slavin & Madden (2004) do note that under SFA reforms, gains can be seen across all academic areas, but stress that this greatly depended on how thoroughly and appropriately the program was implemented in a given school setting. As a result, it is expected that ASP will produce the greatest mathematics gains and there will be no significant difference between AC and SFA on this measure of academic growth.

When it comes to the factors impacting student achievement growth on both the mathematics and reading components of the *TerraNova* test, it is predicted that the variables of race/ethnicity and Community Disadvantage Index (CDI) will produce significant effects, with White students outscoring Black and Hispanic students and those attending schools rating low on the CDI outscoring those attending schools rating high on the CDI. It is also expected that there will be a significant effect of gender on mathematics scores, with boys scoring differently than girls. Lastly, as the CSR programs were designed to produce change in a wide variety of academic settings, it is expected that the school type will not have a significant impact on student growth in mathematics or reading.

CHAPTER THREE

METHOD

Dataset and Sample

Data for this study was taken from the Study of Instructional Improvement (Ball, Cohen, & Rowan, 2010). Data files for this study are accessible through the Inter-University Consortium for Policy and Social Research (ICPSR) website which houses a database of public-use files. No identifying information is included in the files. The primary investigators designed the SII to address the need for a large-scale, longitudinal evaluation of the three most-widely implemented CSR interventions: Accelerated Schools Project, America's Choice, and Success for All.

The researchers collected data from a total of 115 schools from 2000 to 2004. The schools were organized by CSR program: ASP (N = 28), AC (N = 31), and SFA (N = 30); there was also a comparison group of schools not participating in a CSR program (N =26). Schools were chosen purposefully in order to maintain an unbiased and representative sample. Characteristics including school size, poverty level, and demographic diversity were thoughtfully matched for a final sample that was less homogenous than those used in similar studies (Ball, Cohen, & Rowan, 2010). They refer to the Educational Longitudinal Study (ECLS) to emphasize that the SII sample would better reflect the minority experience, as well as that of students coming from a low socio-economic environment, as these are the students who stand to benefit most from CSR interventions.

Data was collected for different cohort groups using a method which involved staggering the introduction of schools into the study. During the first year of data collection, only kindergarten (representing Cohort A) and 3rd grade (representing Cohort B) were tracked; during the second year, Cohort A students were in kindergarten and 1st grade while Cohort B students were in 3rd grade and 4th grade; in the third year Cohort A students were in 1st and 2nd grade while Cohort B students were in 4th and 5th grade; during the fourth and final year, Cohort A 2nd graders and Cohort B 5th graders were monitored. This meant that each student grouping went through three cycles (three school years), with half completing participation in the 2002-2003 school year and the remainder finishing at the end of the 2003-2004 school year. For the purposes of this investigation, data for Cohort A grades 1 and 2 was used for all stages of the analysis; Cohort B grade 4 and 5 data did not have a large enough N to warrant inclusion in the growth model analysis, but was used in the analysis comparing *TerraNova* and *Supera* scores.

Student participants in the SII had a total N of 6,733. Approximately half were male (51.2%) while just under half were African American (49.8%), nearly a quarter were White (23.1%), and approximately one fifth of students participating were Hispanic (19.2%). Only 10.1% of students' mothers had completed college and nearly a quarter of the families earned fewer than $15,000 per annum (24.6%) or received food stamps in the last year (23.0%).

Instruments

TerraNova Test.

The second edition of the *TerraNova* test was used during data collection for the SII. This test was designed to measure students' knowledge of concepts and learning objectives taught in schools across the United States. Reading/Language Arts content was developed using curricular guides from a variety of states and school districts, as well as from parochial school administrations. It is designed to assess students' understanding of and ability to apply critical-thinking and communication skills such as reading comprehension, vocabulary, and language expression. mathematics content was developed to reflect the standards of the National Council of Teachers of mathematics (NCTM), state- and city-wide curriculums, and the achievement goals outlined by the National Assessment of Educational Progress (NAEP). The purpose of the mathematics section of the test is to evaluate students' ability to use Mathematic reasoning (i.e., estimation and computation) for solving real life problems.

TerraNova is a nationally-normed test using a stratified random sample taken from a variety of school types and includes factors such as geographic region, socio-economic status, and race/ethnicity. The second edition of the exam was standardized in 1999 and 2000 (concurrent with SII data collection) using samples of more than 300,000 students. Student performance is reported in several formats: Scale Score, National Percentile, National Stanine, Grade Equivalent, and Normal Curve Equivalent (CTB/McGraw Hill, 2000). For the purposes of this statistical analysis, scale scores for both the mathematics and reading test sections will be used.

***Supera* Test.**

The *Supera* test is the Spanish-Language adaptation of the *TerraNova* test form and is designed to allow educators to assess Spanish-speaking students' academic aptitude using the same scale as the English-language version. *Supera Evaluaciones Múltiples (Multiple Assessments)*, measures Mathematics and Reading/Language Arts content knowledge and features selected-response and constructed response test items.

Variables of Interest

In order to answer the research questions, this study will utilize certain key variables from the full data set (see Table 1 below for all variables that will be included in the analysis). The first goal is to establish if the *TerraNova* and *Supera* tests produce comparable results. For this investigation it will be necessary to observe the relationship between students' mathematics and reading scores on the two tests in the fall and spring for each student cohort. In addition, descriptive student demographic values for gender, race/ethnicity, and Community Disadvantage Index (CDI) for each subsample will be used to enrich the researcher's understanding of the numeric relationship observed. CDI is a composite measure of socio-economic status (SES) which factors in not only income, but also education level, employment, and single-parent status (Ball, Cohen, & Rowan, 2010).

The next aim of this study is to compare student performance on the *TerraNova* test over time based on which CSR intervention was in place within the schools. This will be achieved using *TerraNova* mathematics and reading scores (fall and spring for two academic years) and the school initiative in effect (A, S, or X) [See Table 1 below]. The data will be used to approximate an average growth trend which will then be used to evaluate the performance of the three interventions. The goal is not to demonstrate which CSR program worked *best*, but rather to establish if there were significant differences in impact across the programs.

The third goal of this study is to identify student- and school-level variables which affect the magnitude of impact of the three CSR initiatives. To accomplish this, *TerraNova* mathematics and reading longitudinal gains will be modeled with variables such as school type, CDI, motivation, and student race or gender to determine which factors influence the rate of student growth over time.

Table 1. Variables Included in the Analysis

Variable Name	Variable Type	Variable Range/ Possible Values
General Variables		
School Initiative	categorical	AC, SFA, Comparison, ASP
Student Variables		
Gender	categorical	Male Female
Race/Ethnicity	categorical	Black White Hispanic
Reading Motivation	ordinal	Values 1-4 (4= High and 1= Low)
Math Motivation	ordinal	Values 1-4 (4= High and 1= Low)
School Variables		
School Type	categorical	Charter School Magnet School Regular Elementary
Community Disadvantage Index	ordinal	Values 0-5 (5= High CD and 0= Low CD)
Outcome Variables		
TerraNova Reading Scale Score	continuous	Score Range: 355 – 780
Supera Reading Scale Score	continuous	Score Range: 407 – 722
TerraNova Math Scale Score	continuous	Score Range: 299 – 720
Supera Math Scale Score	continuous	Score Range: 324 – 628

Data Analysis

Descriptive Analysis.

A preliminary descriptive analysis was conducted for Cohort A and Cohort B demographic variables to ensure that there are no significant group differences that might threaten the reliability and validity of subsequent analyses. The researcher compared the cohort groups on variables such as race/ethnicity, gender, and SES (using CDI). Significant sample size differences between the two cohort groups resulted in a reassessment of what data sets were included in the analyses. For the purposes of the present investigation, only Cohort A (1st and 2nd grade) data is used.

Instrument analysis. In order to answer the first research question (Are student outcomes on *TerraNova* mirrored by outcomes on *Supera* (the Spanish language *TerraNova* test)?) the researcher conducted a series of *t*-tests for each group comparing mathematics and reading outcomes for students using the *TerraNova* English-language test form to those of students using the *Supera* Spanish-language version to determine whether the means of two groups are statistically different from each other. This analysis was conducted at the student level. One obstacle to this investigation is the fact that the sample sizes for the two tests groups in Cohort A are unbalanced.

Using a random sample of 81 *TerraNova* scores to serve as the reference group, it will be possible to compare Cohort A pair groups (*TerraNova* Mathematics/*Supera* Mathematics and *TerraNova* Reading/*Supera* Reading) using a series of *t*-tests where the test (*Supera* or *TerraNova*) represents the within subjects variable and time (the four data collection points) represents the between-subjects variable. In addition, a repeated-measures analysis of variance will be conducted to further explore the interrelationship of the time, test form, and achievement variables.

Longitudinal Growth Analysis.

Multi-level modeling will be used to answer the second and third research questions. To address these questions (How do the three CSR programs (ASP, AC, and SFA) compare in terms of how they impact student growth over time based on *TerraNova* test scores? and What are the effects of student- and school-level factors on the outcomes of these comprehensive school reform initiatives?), longitudinal quadratic growth modeling will be used to establish what average student growth in mathematics and reading looks like over time. Using Raudenbush and Bryk's (2002) guidelines, in the first stage, the model will be unconditional containing only the outcomes and time variables. No student-level variables will be included in this stage of analysis. The goal will be to describe average growth produced within each of the subject areas**.** The Level-1 models are

$$Y_{Read\_ij} = \pi_{0i} + \pi_{1i}(\mathrm{TIME})_{ti} + \pi_{2i}(\mathrm{TIME}^2)_{ti} + e_{ti}$$

and

$$Y_{Read\_ij} = \pi_{0i} + \pi_{1i}(\mathrm{TIME})_{ti} + \pi_{2i}(\mathrm{TIME}^2)_{ti} + e_{ti}$$

which represent the average models for the outcomes Y_{Read_ij},the *TerraNova* Reading score, and Y_{Math_ij} ,the *TerraNova* Mathematics score, where π_{0i} is the intercept for person i if $\pi_{1i}(\mathrm{TIME})_{ti} = 0$. The term $\pi_{1i}(\mathrm{TIME})_{ti}$ represents the growth rate for person i for the duration of data collection. The term $\pi_{2i}(\mathrm{TIME}^2)_{ti}$ represents the acceleration at which the change is occurring. The three Level-2 equations which signify the simplest person-level model with no predictors included are

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

$$\pi_{1i} = \beta_{20} + r_{2i}.$$

In the next modeling stage certain demographic predictors will be introduced to the model and checked for significance in order to begin to answer the third research question. The student-level variables gender and race/ethnicity will be added to the model. Individual gender and race variables will be reduced to dichotomous variable types (GENDER: 1 = Male and 0 = Female; HISPANIC: 1 = Hispanic and 0 = Not Hispanic; WHITE: 1 = White and 0 = Not White). Black students will represent the reference group for the race/ethnicity variable and female students will represent the reference group for the gender variable. The new models will be represented by

Level 1: $Y_{Read\_ij} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME}^2)_{ti} + e_{ti}$

Level 2: $\pi_{0i} = \beta_{00} + \beta_{01} * (\text{GENDER})_{\text{i}} + \beta_{02} * (\text{WHITE})_{\text{i}} + \beta_{03} * (\text{HISPANIC})_{\text{i}} + r_{0i}$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (\text{GENDER})_{\text{i}} + \beta_{12} * (\text{WHITE})_{\text{i}} + \beta_{13} * (\text{HISPANIC})_{\text{i}} + r_{1i}$$

$$\pi_{1i} = \beta_{20} + \beta_{21} * (\text{GENDER})_{\text{i}} + \beta_{22} * (\text{WHITE})_{\text{i}} + \beta_{23} * (\text{HISPANIC})_{\text{i}} + r_{2i}.$$

and

Level 1: $Y_{Math\_ij} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME}^2)_{ti} + e_{ti}$

Level 2: $\pi_{0i} = \beta_{00} + \beta_{01} * (\text{GENDER})_{\text{i}} + \beta_{02} * (\text{WHITE})_{\text{i}} + \beta_{03} * (\text{HISPANIC})_{\text{i}} + r_{0i}$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (\text{GENDER})_{\text{i}} + \beta_{12} * (\text{WHITE})_{\text{i}} + \beta_{13} * (\text{HISPANIC})_{\text{i}} + r_{1i}$$

$$\pi_{1i} = \beta_{20} + \beta_{21} * (\text{GENDER})_{\text{i}} + \beta_{22} * (\text{WHITE})_{\text{i}} + \beta_{23} * (\text{HISPANIC})_{\text{i}} + r_{2i}.$$

In the next stage of modeling, after removing the extraneous non-significant demographic variables, the analysis will be run with certain student-level and school-level predictor variables included incrementally. This will include a measure of motivation for mathematics and for reading (low values represent low motivation and high values represent high motivation). (As in the previous model, certain categorical variables will be transformed into dichotomous variables to increase ease of interpretation for the student-level factors and for the school-level factors School Initiative (AC:1 = America's Choice and 0 = Other; ASP: 1= Accelerated Schools Project and 0 = Other; SFA: 1 = Success for All and 0 = Other) and School Type (MAG: 1 = Magnet School and 0 = Other Type of School; CHAR: 1 = Charter School and 0 = Other Type of School). In this model, the reference groups are Black, female students attending regular elementary schools. After all adjustments and removal of non-significant predictors, the final models are represented by

Level 1: $Y_{Read\_ij} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME}^2)_{ti} + e_{ti}$

Level 2: $\pi_{0i} = \beta_{00} + \beta_{01}{*}\,(\text{GENDER})_i + \beta_{02}{*}\,(\text{WHITE})_i + \beta_{03}{*}\,(\text{RMOT})_i + \beta_{04}{*}\,(\text{AC})_i + r_{0i.}$

$\pi_{1i} = \beta_{10} + \beta_{11}{*}\,(\text{GENDER})_i + \beta_{12}{*}\,(\text{WHITE})_i + \beta_{13}{*}\,(\text{RMOT})_i + \beta_{14}{*}\,(\text{AC})_i + r_{1i.}$

$\pi_{1i} = \beta_{20} + \beta_{21}{*}\,(\text{GENDER})_i + \beta_{22}{*}\,(\text{WHITE})_i + \beta_{23}{*}\,(\text{RMOT})_i + \beta_{24}{*}\,(\text{AC})_i + r_{2i}.$

and

Level 1: $Y_{Math\_ij} = \pi_{0i} + \pi_{1i}(\text{TIME})_{ti} + \pi_{2i}(\text{TIME}^2)_{ti} + e_{ti}$

Level 2: $\pi_{0i} = \beta_{00} + \beta_{01}{*}\,(\text{WHITE})_i + \beta_{02}{*}\,(\text{CDI})_i + \beta_{03}{*}\,(\text{AC})_i + \beta_{04}{*}\,(\text{ASP})_i + r_{0i}$

$\pi_{1i} = \beta_{10} + \beta_{11}{*}\,(\text{WHITE})_i + \beta_{12}{*}\,(\text{CDI})_i + \beta_{13}{*}\,(\text{AC})_i + \beta_{14}{*}\,(\text{ASP})_i + r_{1i}$

$\pi_{1i} = \beta_{20} + \beta_{21}{*}\,(\text{WHITE})_i + \beta_{22}{*}\,(\text{CDI})_i + \beta_{23}{*}\,(\text{AC})_i + \beta_{24}{*}\,(\text{ASP})_i + r_{2i}.$

CHAPTER FOUR

RESULTS

Descriptive Results

This study explored student achievement using *TerraNova* test data in an effort to understand the factors that impact student growth. The data are analyzed at the student level. Table 2 contains demographic data for Cohort A students. There was a nearly equal percent of male and female student participants included in the study. In addition, more than three quarters of the sample is non-White. This is atypical for school data where the sample tends to underrepresent students of color and is owed to the deliberate effort of the primary investigators to make the study fully inclusive. In the present analysis, when race/ethnicity is considered, only the three largest demographic groups will be included.

Table 2. Cohort A Demographics

Group	*n*	%
Male	1091	50.3
Female	1076	49.7
Asian	104	4.8
Black	1076	49.7
Hispanic	389	18.1
White	498	23.0
Other	87	4.1
ASP	543	25.1
SFA	505	23.3
AC	660	30.5
Comparison	459	21.2
Total Students	2167[a]	

[a]Race Ns do not add up to total due to missing data for 13 students.

Instrument Comparison Using Independent Samples *t*-tests

In order to address the first research question, are student outcomes on *TerraNova* mirrored by outcomes on *Supera* (the Spanish language *TerraNova* test), independent samples *t*-tests for each pair of scores was conducted. However, before this could be done, some data preparation was necessary. The *Supera* sample contained scores for 81 students, but the *TerraNova* sample contained scores for 2167 students. In order to make a valid comparison between outcomes on the two tests, it was necessary to create a sample of *TerraNova* scores with the same N as the sample of *Supera* scores that would still be representative of the overall trend of *TerraNova* student performance.

***Supera* and *TerraNova* Comparison.**

Using a random sample of 81 *TerraNova* scores, it was possible to conduct the *t*-tests between *TerraNova* and *Supera* scores. Results of these tests are presented in Table 3 and Table 4. While some findings were consistent with the expectations of the researcher, other findings were surprising and raised many questions about how group differences might be accounted for in this, as well as subsequent research.

Reading.

The *t*-tests for reading scores yielded unexpected results. Consistent with expectations, the analysis revealed that for reading there is a significant difference between student scores at the first data collection point (Fall 1), $t(160) = 3.41$, $p = .001$, with students taking the *TerraNova* test scoring higher on average than those taking the *Supera* test. However, the subsequent pairs were not significantly different from each other ($p > .05$). This suggests that students' reading achievement is being measured equivalently on the two tests and that differences are erased after the Fall data collection, but also raises questions about the relationship of the test group to growth over time.

Table 3. Means and Standard Deviations for *Supera* and *TerraNova* Reading Scores by Semester

Group	Fall 1 *M (SD)*	Spring 1 *M (SD)*	Fall 2 *M (SD)*	Spring 2 *M (SD)*
All Students	511.66 (49.29)	535.62 (43.64)	567.58 (38.48)	580.15 (35.41)
Supera	498.88 (51.78)*	541.58 (48.54)	562.61 (38.73)	581.21 (33.92)
TerraNova	524.44 (43.33)*	531.74 (40.02)	570.86 (38.34)	579.63 (36.45)
Number Tested				
All Students	162	114	83	73
Supera	81	45	33	25
TerraNova	81	69	50	49

*Significant group differences at the p<.01 level.

Mathematics.

Mathematics score *t*-tests also revealed unique differences between the two test groups. Unlike reading scores, which were not significantly different after the Fall 1 data collection, *TerraNova* and *Supera* mathematics scores were significantly different from each other at all four data collection points. At Fall 1, students taking *TerraNova* outscore students taking *Supera* by nearly 50 points, $t(160) = 7.29$, $p < .001$. While this gap does close gradually from Fall 1 to Spring 2, there are still significant differences between all score pairs. It is worth noting that the Spring 2 scores approached non-significance, $t(72) = 2.02$, $p = .047$. As with reading scores, this poses new questions about the relationship between time and academic growth for students in these two test groups.

Table 4. Means and Standard Deviations for *Supera* and *TerraNova* Mathematics Scores by Semester

Group	Fall 1 *M (SD)*	Spring 1 *M (SD)*	Fall 2 *M (SD)*	Spring 2 *M (SD)*
All Students	448.24 (47.88)	485.10 (40.02)	507.24 (37.57)	525.01 (44.84)
Supera	424.41 (46.78)**	464.53 (39.73)**	489.70 (35.77)**	510.60 (43.87)*
TerraNova	472.07 (35.44)**	498.51 (34.31)**	518.82 (34.39)**	532.37 (43.95)*
Number Tested				
All Students	162	114	83	74
Supera	81	45	33	25
TerraNova	81	69	50	49

*Significant group differences at the p<.05 level.
**Significant group differences at the p<.001 level.

Follow-up Repeated-Measures Analysis of Variance

In order to further explore the differences between *TerraNova* and *Supera* students' scores, a follow-up repeated-measures ANOVA was conducted for reading and mathematics scores with Time as the within-subjects variable and Test (*Supera* and *TerraNova*) as the between-subjects variable. Results for this analysis are presented in Table 7. Since the assumption of sphericity was violated (Mauchley's Test had a p-value <.05 for both the reading and mathematics ANOVAs), the Greenhouse-Geisser correction was used. As a result, the mean-squares, degrees of freedom, and F values for reading and mathematics outcomes contained in Table 7 represent this adjustment.

As would be expected, there was a significant effect of Time on average performance for reading, $F(1, 2.55) = 56.73$, $p < .001$, and for mathematics, $F(1, 2.54) = 94.38$, $p < .001$, which indicated that over time, there was a change in students' performance in reading and mathematics. What were unexpected were the results for the

Time x Test interaction. For reading scores, there was a significant interaction between Time and Test, $F(1, 2.55) = 7.99$, $p < .001$. This reflected the fact that student growth in reading and mathematics was different depending on which test form, *Supera* or *TerraNova*, was used. However, this was not seen in mathematics scores, $F(1, 2.54) = 1.55$, $p = .210$. Figures 1 and 2 are mean plots for reading and mathematics scores and graphically present the interaction effects (or lack thereof). There is an unusual growth pattern for reading scores, but mathematics achievement follows a linear trend. While it is outside of the scope of this study, further research might shed some light on this phenomenon.

Table 5. Repeated-Measures ANOVA using the Greenhouse-Geisser Sphericity Correction

	MS	*df*	*F*	*p*
Reading				
Time	61009.995	2.55	56.73	<.001
Time x Test	8589.946	2.55	7.99	<.001
Error	1075.445	160.40		
Mathematics				
Time	54229.669	2.54	94.38	<.001
Time x Test	890.42	2.54	1.55	.210
Error	1459.356	64.00		

Figure 1. Mean Plot for Reading Score Time x Test Interaction

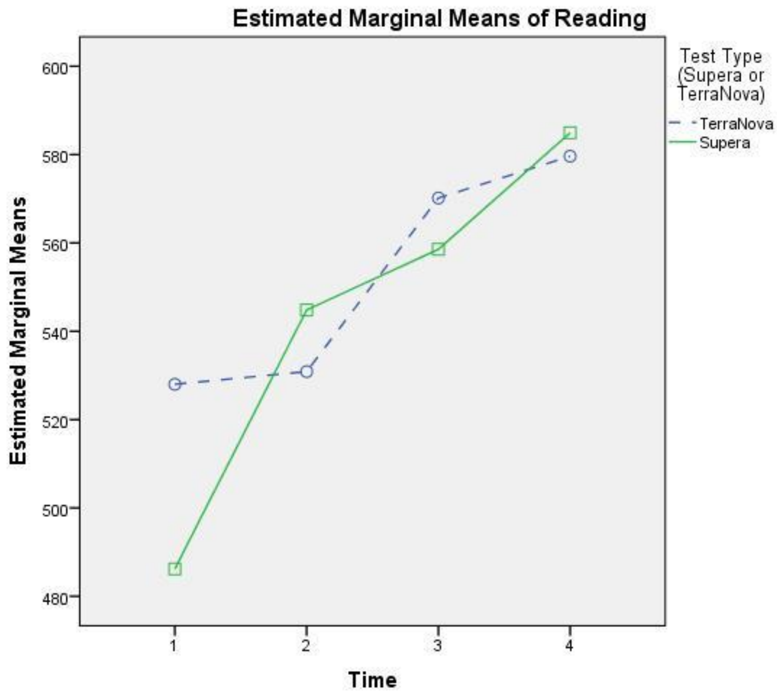
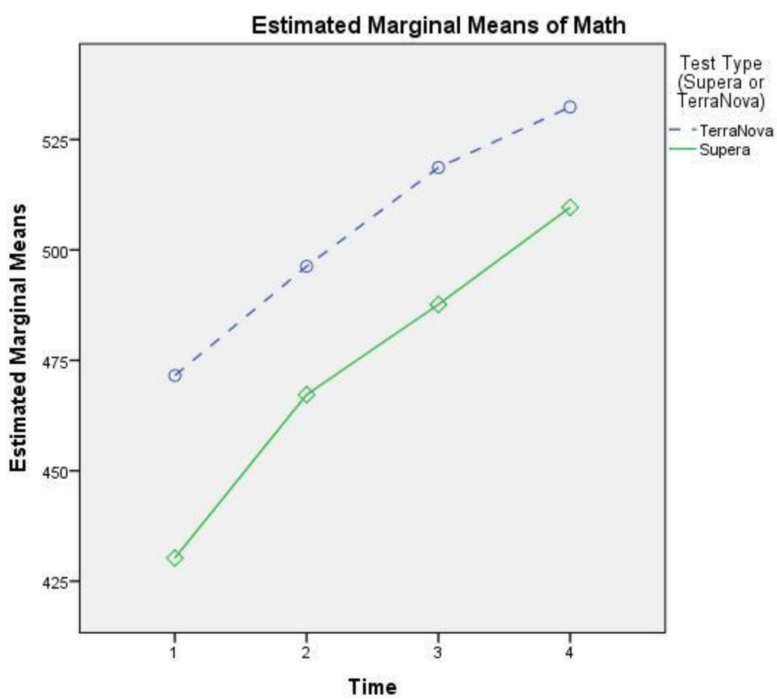


Figure 2. Mean Plot for Mathematics Score Time x Test Interaction



Quadratic Growth Modeling

In order to answer the second and third research questions, hierarchical modeling techniques were applied to student test data, with reading and mathematics scores as the outcome. Using Raudenbush and Bryk's (2002) guidelines, the first stage involved the construction of an unconditional model of student growth over time that includes only the Level-1 time predictors and no student-level variables. In the next stage, certain demographic and group characteristics were added sequentially to the model. Significant predictors were kept in the model and non-significant predictors were removed, as non-significance indicates that there are no differences between students in those groups. As more predictors were added to the model, certain significant predictors lose significance and were removed from subsequent models. The ultimate goal was to arrive at a model that contains only significant predictors of initial achievement and/or the student growth rate over time.

Linear vs. Quadratic.

Initially, a linear growth model was used to explore the academic growth trends of students in reading and mathematics on the *TerraNova* test. However, after running the unconditional model, it became evident that a quadratic model might fit the data better. Graphing the individual growth curves for students' reading and mathematics scores revealed a curved growth trend rather than a linear one. This confirmed that the use of a quadratic growth model would be more suited to the analysis. Examples of growth curves for reading and mathematics can be seen in Figures 3-6. While linear growth modeling uses the time variable as is, quadratic growth modeling uses time squared, which provides an *accelerated* growth variable. The purpose is to show that change over time does not occur at a consistent rate as it does with linear data; rather, the rate of growth itself changes at a constant rate (Raudenbush & Bryk, 2002).

Figure 3. Curvilinear Growth Trend for One Student using Reading Scale Score (RSS)

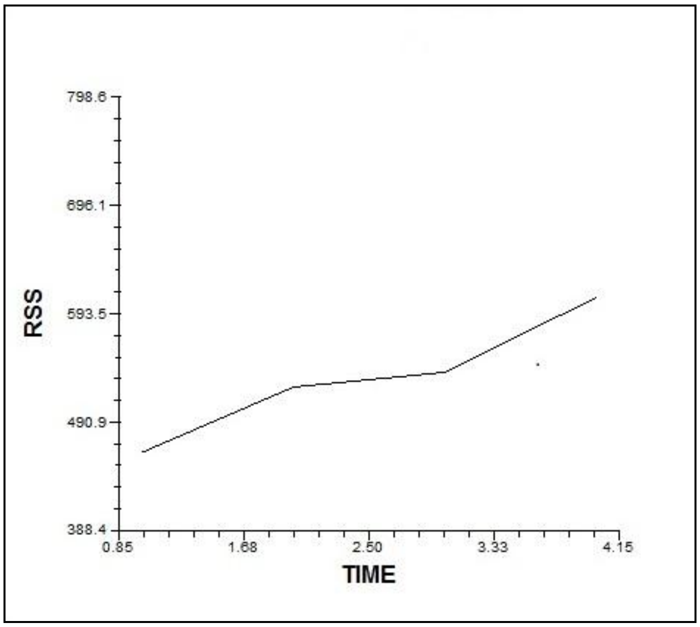


Figure 4. Curvilinear Growth Trend for One Student using Reading Scale Score (RSS)

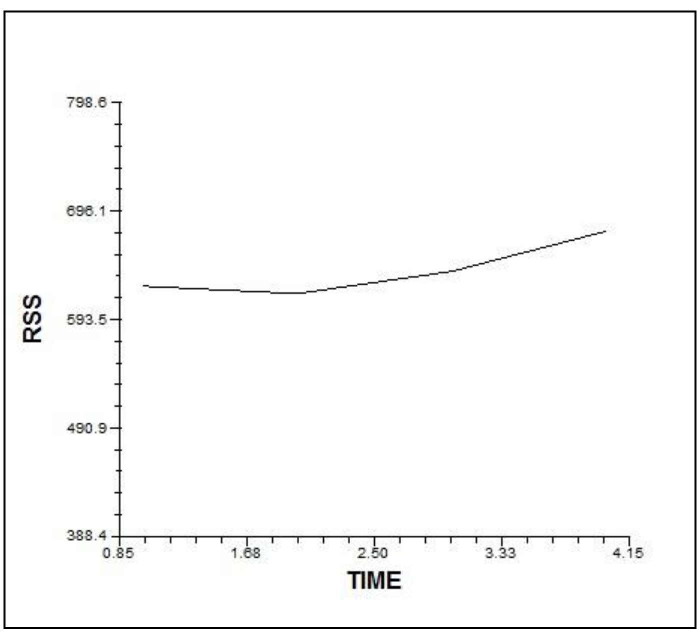


Figure 5. Curvilinear Growth Trend for One Student using Mathematics Scale Score (MSS)

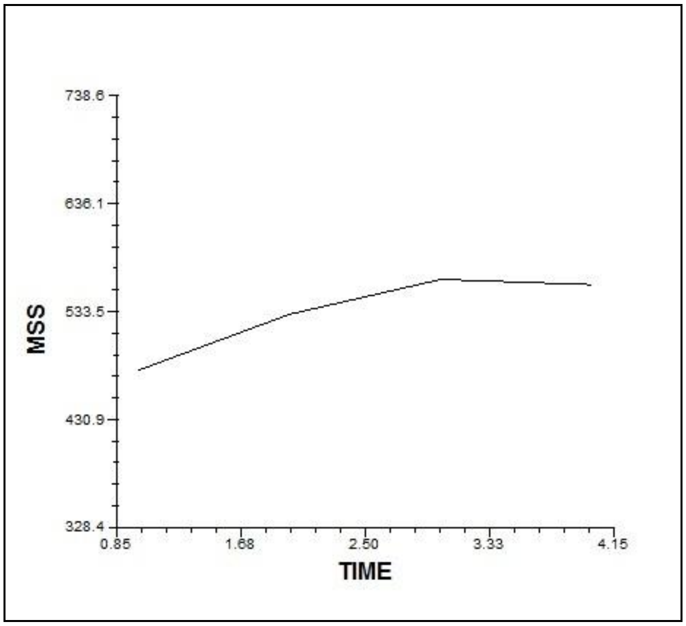
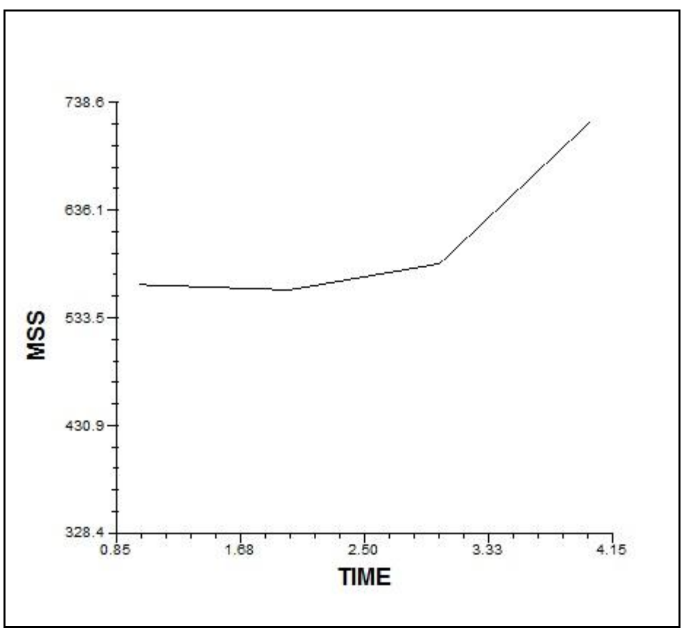


Figure 6. Curvilinear Growth Trend for One Student using Mathematics Scale Score (MSS)



Reading.

Unconditional reading model. Based on the results of the unconditional model run for reading scores, which describes average achievement growth for the average student, the grand means are significantly different from zero (see Tables 6 and 7). In addition, there is significant variation between students at the intercept and the slope. The average initial reading score is 501.33 and there is significant variation around this score, π = 1312.11, $p < .001$. The average linear growth rate is 32.80 points per semester and there is significant variation around the growth rate, π = 1076.01, $p < .001$. The accelerated (quadratic) growth rate is -4.51 points per semester and there is significant variation around the accelerated rate, π = 35.24, $p < .001$. This suggests that over time, reading achievement growth slows by an average of four and half points per semester.

Table 6. Reading Results for the Unconditional Model

Level-1	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	501.33	1.93	259.51	1949	0.000
Time					
Intercept	42.30	1.79	23.58	1949	0.000
Time^2					
Intercept	-4.51	.35	-12.78	1949	0.000

Table 7. Variance Components for the Unconditional Model

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	36.22	1312.11	1506	1794.80	0.000
Time Slope	32.80	1076.01	1506	1795.08	0.000
Time^2 Slope	5.94	35.24	1506	1744.36	0.000
Level-1 Error	26.56	705.30			

Conditional Reading Models.

Gender and race as predictors. The first conditional quadratic growth model includes the demographic characteristics of gender, White, Hispanic, and Black as student-level (Level-2) predictors. Tables 8 and 9 show that the average initial reading score for the average female student is 505.44 and there is significant variation around this initial score, $\pi = 1174.13$, $p < .001$. Over time the accelerated growth rate for the average female student is -4.05 points per semester, indicating that over time average student growth slows about four points each semester, and there is significant variation around the accelerated growth rate, $\pi = 35.95$, $p < .001$.

Initially, the average reading score for male students is 12.66 points lower than the average female student. Over time the accelerated growth rate of the average male student's reading score is 1.60 points slower than the average female student per semester. When it comes to students' race, White, Black, and Hispanic are added to the model to emphasize differences between the three groups. The choice to include and then remove reference groups from the model was made in order to show the specific impacts of different group membership. When it comes to students' race, on average, White students are at a 23.57 point advantage for initial reading score, while Black students are at a 1.36 point disadvantage, and Hispanic students are at a 12.57 point disadvantage, compared to students of other races/ethnicities (Asian, Multi-Ethnic, and Other). Over time, the accelerated growth rate increases by an average of 1.77 points for White students and 0.21 points for Black students, but decreases by 0.82 for Hispanic students. However, these changes are not significant, indicating that race is not the most valuable predictor for reading achievement growth. It may be that including all three categories led to an over-specified model for reading achievement when it came to race, but ultimately only White race status was significant as a predictor of initial status.

Once the variables of gender and race are added to the model the variance is reduced from 1312.11 to 1174.13. Ten percent (10.5%) of the variation in reading scores among students is accounted for by introducing gender and race to the model. Variation among accelerated growth rates in reading scores is still significant and is explained more so by gender differences than race. Introducing gender and race into the model did not reduce the accelerated growth variance, a further indication that the race categories do not predict growth. White status, while not a significant predictor of growth at this stage, was significant as a predictor of initial status and was left in the model. The minority race categories were not significant predictors of initial achievement or growth and are removed from subsequent models. Ultimately, White status is the most significant race predictor in the model.

Table 8. Reading Results for the Conditional Model – Gender and Race as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	505.44	6.35	79.56	1945	0.000
Gender	-12.66	3.82	-3.32	1945	0.001
White	23.57	7.29	3.21	1945	0.002
Black	-1.36	6.69	-0.20	1945	0.839
Hispanic	-12.57	7.49	-1.68	1945	0.093
Time					
Intercept	40.28	5.92	6.80	1945	0.000
Gender	7.54	3.59	2.10	1945	0.035
White	-7.86	6.81	-1.15	1945	0.249
Black	-1.81	6.24	-0.29	1945	0.772
Hispanic	4.64	7.02	0.66	1945	0.509
Time^2					
Intercept	-4.05	1.16	-3.50	1945	0.001
Gender	-1.60	0.71	-2.27	1945	0.023
White	1.77	1.33	1.33	1945	0.184
Black	0.21	1.22	0.18	1945	0.861
Hispanic	-0.82	1.38	-0.60	1945	0.551

Table 9. Variance Components for the Conditional Model – Gender and Race as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	34.27	1174.13	1502	1754.17	0.000
Time Slope	33.10	1095.35	1502	1797.55	0.000
Time^2 Slope	6.00	35.95	1502	1743.43	0.000
Level-1Error	26.46	700.34			

Gender, race, motivation, and Community Disadvantage Index (CDI) as predictors. In the next stage of modeling, after the non-significant race variables were removed, students' reading motivation score and the CDI score were introduced into the model. Tables 10 and 11 show that the average initial reading score for the average female student of color with average reading motivation and average community disadvantage is 501.89 and there is significant variation around this initial score, π =

1151.59, $p < .001$. Over time, the accelerated growth rate for the average student is -4.15 points per semester while controlling for the gender, ethnicity, motivation, and the community disadvantage. It indicates that over time average student growth slows about four points each semester. The variation around the accelerated growth rate is significant, $\pi = 33.70$, $p < .001$.

Initially, the average reading score for male students is 12.27 points lower than the average female student. Over time the accelerated growth rate of the average male student's reading score is 1.88 points slower than the average female student each semester. When it comes to students' race, on average, White students are at a 24.71 point advantage for initial reading score compared to students of color. Over time, the accelerated growth rate increases by an average of 2.56 points for White students.

Results indicate that reading motivation was positively related to reading achievement ($\beta = 8.84$, $p = .020$) while controlling for gender, race, and CDI, which suggests that students more highly motivated in reading will tend to score higher than those with low motivation. CDI was negatively related to initial reading score ($\beta = - 1.05$) while controlling for gender, race, and motivation, but this was not significant, $p = .397$. In addition, there are significant differences in the accelerated growth rates based on motivation and CDI. Student motivation is associated with an accelerated growth rate of -2.69 points per semester, while CDI is associated with a 0.22 point per semester increase, but this is not significant ($p = .327$) and is not included in subsequent reading models.

Once the variables of motivation and CDI are added to the model the variance is reduced from 1312.11 to 1151.59. Twelve percent (12.2%) of the variation in reading scores among students is accounted for by introducing motivation and CDI to the model. At this stage, variation among accelerated growth rates in reading scores is explained by gender, race, and motivation and is still significant.

Table 10. Reading Results for the Conditional Model – Gender, Race, Motivation, and CDI as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	501.89	2.90	172.80	1945	0.000
Gender	-12.27	3.83	-3.21	1945	0.002
White	24.71	4.92	5.02	1945	0.000
Motivation	8.84	3.79	2.33	1945	0.020
CDI	-1.05	1.24	-0.85	1945	0.397
Time					
Intercept	40.45	2.71	14.90	1945	0.000
Gender	8.94	3.58	2.50	1945	0.013
White	-11.75	4.60	-2.44	1945	0.011
Motivation	13.37	3.55	3.76	1945	0.000
CDI	-1.24	1.16	-1.07	1945	0.284
Time^2					
Intercept	-4.15	0.53	-7.79	1945	0.000
Gender	-1.88	0.70	-2.67	1945	0.008
White	2.56	0.91	2.83	1945	0.005
Motivation	-2.69	0.70	-3.85	1945	0.000
CDI	0.22	0.23	0.98	1945	0.327

Table 11. Variance Components for the Conditional Model – Gender, Race, Motivation, and CDI as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	33.94	1151.59	1502	1744.30	0.000
Time Slope	32.21	1037.61	1502	1777.50	0.000
Time^2 Slope	5.81	33.70	1502	1724.38	0.000
Level-1Error	26.48	701.01			

Gender, race, motivation, and school type as predictors. In the next stage, the type of school students attend is added as predictor to the level-two model. As can be seen in Tables 12 and 13, gender, race, and reading motivation are still significant as predictors of initial reading achievement, as well as growth. However, no significant differences were found between students in different types of schools and they were removed from the next and final models. Nonetheless, introducing these variables reduced error variance from 1312.11 to 1135.33 (13.5% explained by the predictors). As with race in the first model, the choice was made to include all three school types, Regular, Magnet, and Charter, in this model. Intermediate modeling stages looked at the model with the reference group (Regular) left out of the model and also included and no differences in significance of the other school types were found. School type was not found to be a relevant predictor for students' reading scores.

Table 12. Reading Results for the Conditional Model – Gender, Race, Motivation, and School Type as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	507.54	9.82	51.69	1943	0.000
Gender	-12.07	3.82	-3.16	1943	0.002
White	25.94	4.58	5.66	1943	0.000
Motivation	9.02	3.79	2.38	1943	0.017
Regular	-7.68	9.77	-0.79	1943	0.432
Magnet	4.36	10.89	0.40	1943	0.688
Charter	-32.55	21.06	-1.55	1943	0.122
Time					
Intercept	46.96	9.19	5.11	1943	0.000
Gender	8.91	3.58	2.49	1943	0.013
White	-9.78	4.29	-2.28	1943	0.023
Motivation	13.50	3.56	3.80	1943	0.000
Regular	-6.57	9.15	-0.72	1943	0.473
Magnet	-13.81	10.19	-1.36	1943	0.176
Charter	23.93	19.71	1.21	1943	0.225
Time^2					
Intercept	-5.61	1.80	-3.11	1943	0.002
Gender	-1.88	0.70	-2.68	1943	0.008
White	2.21	0.84	2.62	1943	0.009
Motivation	-2.73	0.70	-3.90	1943	0.000
Regular	1.54	1.80	0.86	1943	0.391
Magnet	2.59	2.00	1.29	1943	0.196
Charter	-4.55	3.88	-1.17	1943	0.242

Table 13. Variance Components for the Conditional Model – Gender, Race, Motivation, and School Type as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	33.69	1135.33	1500	1738.69	0.000
Time Slope	32.09	1029.55	1500	1772.06	0.000
Time^2 Slope	5.78	33.40	1500	1719.24	0.000
Level-1Error	26.48	701.11			

Gender, race, motivation, and CSR program as predictors. In the next phase, after the non-significant variables were removed, the CSR programs (AC, SFA, and ASP) were introduced into the model, with the control schools serving as the reference group. Tables 14 and 15 show that the average initial reading score for the average female student of color with average reading motivation not attending a CSR program school is 498.06 and there is significant variation around this initial score, $\pi = 1098.47$, $p < .001$. Over time, the accelerated growth rate for the average student is -5.35 points per semester, indicating that over time average student growth slows almost five and half points each semester, and there is significant variation around the accelerated growth rate, $\pi = 31.93$, $p < .001$.

Initially, the average reading score for male students is 11.73 points lower than the average female student. Over time the accelerated growth rate of the average male student's reading score is 1.83 points slower than the average female student per semester. When it comes to students' race, on average, White students are at a 29.08 point advantage for initial reading score compared to students of color. Over time, the accelerated growth rate increases by an average of 2.56 points for White students per semester. Motivation is positively related to initial status ($\beta = 8.99$, $p = .017$) and negatively related to the accelerated growth rate ($\beta = -2.70$, $p < .001$). Results indicate that students in an AC school have initial reading scores which are 14.58 points higher on average than other students. Students in SFA and ASP schools did not differ significantly from other students' initial reading status ($p > .05$).

There are significant differences in the accelerated growth rates for students based on CSR program implementation. Being in an AC school is associated with an accelerated growth rate of 3.18 points per semester. Students in SFA or ASP schools did not differ significantly from other students in growth rate. In the next model, SFA was removed and ASP was been retained because SFA has a lower coefficient value (β = -0.85, p = .418) than ASP (β = 1.59, p = .118). As a result, the accelerated growth rate for ASP students becomes significant (p = .008). However, this also introduced error into the model; the variance grew from 1312.11 in the unconditional model to 1883.18 (+43.5%) and so it was also removed from the final model.

Once the CSR program groups were added to the model the variance was reduced from 1312.11 to 1098.47. More than sixteen percent (16.3%) of the variation in reading scores among students was accounted for by adding these program groups to the model. At this stage, variation among accelerated growth rates in reading scores is explained by gender, race, motivation, and CSR program participation and is still significant.

Table 14. Reading Results for the Conditional Model – Gender, Race, Motivation, and CSR Program as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	498.06	4.47	111.35	1942	0.000
Gender	-11.73	3.81	-3.08	1942	0.003
White	29.08	4.68	6.21	1942	0.000
Motivation	8.99	3.77	2.38	1942	0.017
AC	14.58	5.31	2.75	1942	0.007
SFA	-10.33	5.73	-1.80	1942	0.071
ASP	1.51	5.53	0.27	1942	0.785
Time					
Intercept	46.69	4.18	11.18	1942	0.000
Gender	8.60	3.57	2.41	1942	0.016
White	-12.05	4.37	-2.76	1942	0.006
Motivation	13.34	3.53	3.77	1942	0.000
AC	-17.84	4.96	-3.60	1942	0.001
SFA	6.05	5.36	1.13	1942	0.260
ASP	-7.42	5.18	-1.43	1942	0.152
Time^2					
Intercept	-5.35	0.82	-6.52	1942	0.000
Gender	-1.83	0.70	-2.62	1942	0.009
White	2.56	0.86	2.97	1942	0.003
Motivation	-2.70	0.70	-3.87	1942	0.000
AC	3.18	0.98	3.26	1942	0.001
SFA	-0.85	1.06	-0.81	1942	0.418
ASP	1.59	1.02	1.56	1942	0.118

Table 15. Variance Components for the Conditional Model – Gender, Race, Motivation, and CSR Program as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	33.14	1098.47	1500	1734.25	0.000
Time Slope	31.28	978.28	1500	1760.20	0.000
Time^2 Slope	5.65	31.93	1500	1710.51	0.000
Level-1Error	26.46	70.33			

Final model. In the final model, after the non-significant variables were removed, the intercept represents the average reading achievement for female students of color with average reading motivation who did not attend an AC school. As is shown in Tables 16 and 17, the average initial reading score for the average female student of color with average reading motivation not attending an AC school is 495.41 and there is significant variation around this initial score, $\pi = 1107.51$, $p < .001$. Over time, the accelerated growth rate for the average student is -5.09 points per semester, indicating that over time average student accelerated growth slows about five points, and there is significant variation around this rate, $\pi = 32.42$, $p < .001$.

Initially, the average reading score for male students is 11.98 points lower than the average female student. Over time the accelerated growth rate of the average male student's reading score is 1.84 points slower than the average female student per semester. When it comes to students' race, on average, White students are at a 29.58 point advantage for initial reading score compared to students of color. Over time, the accelerated growth rate increases by an average of 2.80 points for White students per semester. Motivation is still positively related to initial status ($\beta = 8.90$, $p = .019$) and negatively related to accelerated growth ($\beta = -2.70$, $p < .001$) when controlling for gender and race; this indicates that students highly motivated in reading are more likely to have a higher initial reading score, but over time their growth rate will slow about two and a half points per semester compared to students who were not as highly motivated. Results indicate that students in an AC school have initial reading scores which are 17.28 points higher on average than other students and being in an AC school is associated with an accelerated growth rate of 2.90 points per semester.

In the final reading model the variance is reduced from 1312.11 to 1107.51. Approximately fifteen and half percent (15.6%) of the variation in reading scores among students is accounted for by students' gender, race, reading motivation, and attendance at an AC school. At this stage, variation among accelerated growth rates in reading scores is explained by gender, race, motivation, and AC school attendance and is still significant. While this model contains no non-significant predictors and error variance has been reduced from the unconditional model, a great deal of variance in student reading scores is still left unexplained. This suggests that there are other variables not included that account for differential achievement and growth rate.

Table 16. Reading Results for the Final Conditional Model – Gender, Race, Motivation, and CSR Program as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	495.41	3.23	153.45	1945	0.000
Gender	-11.98	3.81	-3.14	1945	0.002
White	29.58	4.63	6.39	1945	0.000
Motivation	8.90	3.77	2.36	1945	0.019
AC	17.28	4.21	4.10	1945	0.000
Time					
Intercept	46.09	3.02	15.27	1945	0.000
Gender	8.71	3.57	2.44	1945	0.015
White	-13.26	4.33	-3.06	1945	0.003
Motivation	13.38	3.54	3.78	1945	0.000
AC	-17.16	3.93	-4.36	1945	0.000
Time^2					
Intercept	-5.09	0.59	-8.58	1945	0.000
Gender	-1.84	0.70	-2.63	1945	0.009
White	2.80	0.85	3.29	1945	0.001
Motivation	-2.70	0.70	-3.87	1945	0.000
AC	2.90	0.77	3.74	1945	0.000

Table 17. Variance Components for the Final Conditional Model – Gender, Race, Motivation, and CSR Program as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	33.28	1107.51	1502	1737.93	0.000
Time Slope	31.50	992.42	1502	1764.98	0.000
$Time^2$ Slope	5.69	32.42	1502	1715.05	0.000
Level-1Error	26.47	700.82			

Mathematics.

Unconditional mathematics model. Based on the results of the unconditional model run for mathematics scores, which describes average achievement growth for the average student, the grand means are significantly different from zero (see Tables 18 and 19). There is significant variation between students at the intercept and the slope. The average initial mathematics score is 463.97 and there is significant variation around this score, $\pi = 2422.35$, $p < .001$. The average linear growth rate is 30.31 points per semester and there is significant variation around the growth rate, $\pi = 830.56$, $p < .001$. The accelerated (quadratic) growth rate is -1.63 points per semester and there is significant variation around the accelerated rate, $\pi = 35.42$, $p < .001$. This suggests that over time, students' math achievement growth slows by an average of more than a point and a half per semester.

Table 18. Mathematics Results for the Unconditional Model

Level-1	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	463.97	1.77	261.61	1949	0.000
Time					
Intercept	30.31	1.45	20.92	1949	0.000
$Time^2$					
Intercept	-1.63	0.29	-5.54	1949	0.000

Table 19. Variance Components for the Unconditional Model

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	49.22	2422.35	1505	2389.13	0.000
Time Slope	28.82	830.56	1505	1835.01	0.000
Time^2 Slope	5.95	35.42	1505	1881.25	0.000
Level-1Error	20.88	435.90			

Conditional Mathematics Models.

Gender and race as predictors. The first conditional quadratic growth model for mathematics achievement includes the demographic characteristics of gender, White, Hispanic, and Black as student-level (Level-2) predictors. Tables 20 and 21 show that the average initial mathematics score for the average female student is 467.32 and there is significant variation around this initial score, $\pi = 1915.04$, $p < .001$. The accelerated growth rate for the average female student is -0.55 points per semester, indicating that over time average student growth slows about half a point each semester, and there is significant variation around the accelerated growth rate, $\pi = 29.84$, $p < .001$, but this does not reflect a significant change, $p = .562$.

Initially, the average mathematics score for male students is 0.85 points lower than the average female student. Over time the accelerated growth rate of the average male student's reading score is 0.15 points slower than the average female student per semester. As in the reading model, when it comes to students' race, three race categories, White, Black, and Hispanic are added to the model to highlight the disparities in achievement among the three groups. On average, White students are at a 35.92 point advantage for initial mathematics score, while Black students are at a 15.29 point disadvantage, and Hispanic students are at a 19.57 point disadvantage. Over time, the accelerated growth rate increases by an average of 2.66 points for White students, but decreases by 2.31 points for Black students and 2.55 points per semester for Hispanic students, compared to students of other races/ethnicities (Asian, Multi-Ethnic, and Other). Though including all these groups increased the risk of over-specifying of the model, it was important to present both the advantage of being a White student, as well as the disadvantage associated with being Hispanic or Black. As with reading, the choice to include and then remove reference groups from the model was made in order to show the specific impacts of different group membership. At this stage, there were no significant differences found between boys and girls when it comes to initial mathematics score or accelerated growth; the gender variable will not be included in subsequent mathematics models.

Once the variables of gender and race are added to the model the variance is reduced from 2422.35 to 1915.04. This means that twenty percent (20.9%) of the variation in mathematics scores among students is explained by introducing gender and race to the model. Variation among accelerated growth rates in mathematics scores is still significant and is explained more so by race than by gender. The coefficient values for the gender and race predictors, paired with *p*-values show that race explains more variation. Introducing gender and race into the model for mathematics achievement reduced the accelerated growth variance from 35.42 to 29.84, meaning that these predictors explain almost sixteen percent (15.8%) of the variance in the accelerated slope.

Table 20. Mathematics Results for the Conditional Model – Gender and Race as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	467.32	5.67	82.37	1945	0.000
Gender	-0.85	3.41	-0.25	1945	0.804
White	35.92	6.50	5.52	1945	0.000
Black	-15.29	5.97	-2.56	1945	0.011
Hispanic	-19.57	6.68	-2.93	1945	0.004
Time					
Intercept	25.43	4.71	5.40	1945	0.000
Gender	0.93	2.86	0.33	1945	0.744
White	-13.07	5.42	-2.41	1945	0.016
Black	9.98	4.97	2.01	1945	0.044
Hispanic	13.34	5.59	2.39	1945	0.017
Time^2					
Intercept	-0.55	0.95	-0.58	1945	0.562
Gender	-0.15	0.58	-0.26	1945	0.793
White	2.66	1.09	2.43	1945	0.015
Black	-2.31	1.00	-2.30	1945	0.021
Hispanic	-2.55	1.13	-2.26	1945	0.024

Table 21. Variance Components for the Conditional Model – Gender and Race as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	43.76	1915.04	1501	2162.03	0.000
Time Slope	26.48	701.12	1501	1773.94	0.000
Time^2 Slope	5.46	29.84	1501	1813.96	0.000
Level-1Error	20.93	437.92			

Race, motivation, and Community Disadvantage Index (CDI) as predictors. The next quadratic growth model for mathematics achievement includes the demographic characteristics of White, Hispanic, and Black, mathematics motivation and CDI as predictors. Tables 22 and 23 show that the average initial mathematics score for the average student is 466.65 and there is significant variation around this initial score, π =

1806.16, $p < .001$. Over time the accelerated growth rate for the average student is -0.73 points per semester, indicating that over time average student growth slows about three quarters of a point, and there is significant variation around the accelerated growth rate, $\pi = 29.84$, $p < .001$. This still is not a significant change.

Initially, the average mathematics score for White students is 29.60 points higher than the average non-White student. Over time the accelerated growth rate for the average White student's score increases by 2.25 points per semester. For initial mathematics scores, Black students are at a 13.43 point disadvantage, and Hispanic students are at a 15.51 point disadvantage, compared to students of other races/ethnicities (Asian, Multi-Ethnic, and Other). Over time, the accelerated growth decreases by an average of 2.06 points for Black students and 2.24 points per semester for Hispanic students. For Hispanic students, initial status differences were statistically significant and accelerated growth approached significance ($p = .049$) and is retained in the next model.

When it came to mathematics motivation, there was a positive relationship to initial status observed ($\beta = 18.28$, $p < .001$) while controlling for race and community disadvantage, such that higher motivation is likely to predict a higher initial mathematics score. As with reading, CDI was negatively related to initial mathematics score ($\beta = -4.05$, $p < .001$) when controlling for race and motivation, indicating that students who live in a more disadvantaged community will not score as highly on the mathematics test as those in areas experiencing less need. In addition, there are significant differences in the accelerated growth rates based on CDI, but not motivation. CDI is associated with an accelerated growth rate of 0.52 points lost per semester, while student motivation is associated with a 0.91 point per semester increase, but is not significant (p = .120) and is not included in subsequent models.

Once the variables of motivation and CDI are added to the model the variance is reduced from 2422.35 to 1806.16. Twenty-five percent (25.4%) of the variation in mathematics scores is explained by bringing motivation and CDI to the model. Variation among accelerated growth rates in mathematics scores among students is still significant and is explained more so by race and CDI than motivation. Introducing motivation into the model also reduced the accelerated growth variance from 35.42 to 29.68, meaning that these predictors explain over sixteen percent (16.2%) of the variance in the slope.

Table 22. Mathematics Results for the Conditional Model – Race, Motivation, and CDI as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	466.65	5.39	86.54	1944	0.000
White	29.60	6.53	4.53	1944	0.000
Black	-13.43	5.94	-2.26	1944	0.024
Hispanic	-15.51	6.66	-2.33	1944	0.020
Motivation	18.28	3.40	5.37	1944	0.000
CDI	-4.05	1.10	-3.67	1944	0.000
Time					
Intercept	26.27	4.52	5.81	1944	0.000
White	-11.55	5.49	-2.10	1944	0.035
Black	9.04	4.99	1.81	1944	0.070
Hispanic	12.09	5.62	2.15	1944	0.031
Motivation	5.15	2.89	1.78	1944	0.074
CDI	2.16	0.94	2.30	1944	0.021
Time^2					
Intercept	-0.73	0.91	-0.80	1944	0.425
White	2.25	1.11	2.03	1944	0.042
Black	-2.06	1.00	-2.05	1944	0.040
Hispanic	-2.24	1.14	-1.97	1944	0.049
Motivation	-0.91	0.59	-1.55	1944	0.120
CDI	-0.52	0.19	-2.74	1944	0.007

Table 23. Variance Components for the Conditional Model – Race, Motivation, and CDI as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	42.50	1806.16	1500	2126.28	0.000
Time Slope	26.45	699.45	1500	1775.57	0.000
Time^2 Slope	5.45	29.68	1500	1814.12	0.000
Level-1Error	20.90	437.01			

Race, CDI, and School Type as predictors. In the next phase of modeling for mathematics achievement, non-significant variables have been removed and the types of schools students attend have been added as predictors. As is shown in Tables 24 and 25, the average initial mathematics score for the average student, regardless of gender is 489.74 and there is significant variation around this initial score, $\pi = 1859.32$, $p < .001$. Over time the accelerated growth rate for the average student is 2.08 points per semester, indicating that over time average student growth increases about 2 points per semester when the school type is considered, and there is significant variation around the accelerated growth rate, $\pi = 29.35$, $p < .001$.

Initially, the average mathematics score for White students is 32.61 points higher than the average non-White student. Over time the accelerated growth rate for the average White student's score increases by 2.21 points per semester. For initial mathematics scores, Black students are at a 12.34 point disadvantage, and Hispanic students are at a 13.87 point disadvantage. Over time, the accelerated growth decreases by an average of 2.00 points for Black students and 1.94 points per semester for Hispanic students. For Hispanic students, initial status differences were still statistically significant, but accelerated growth is no longer approaching significance and is not retained in the next model. The effects of White and Black race statuses on students accelerated mathematics growth rates are approaching significance and will be kept in the next model (p = .046 for both categories).

For CDI as a predictor, there was still a negative relationship to initial status observed (β = -4.39, p < .001) while controlling for race and school type, such that higher CDI (more disadvantaged) is associated with lower initial mathematics scores. In addition, there are significant differences in the accelerated growth rates based on CDI. CDI is associated with an accelerated growth rate of -0.53points per semester. Students from regular schools are at a 26.59 point disadvantage compared to students in other types of schools, and the quadratic growth rate for these students slows about 3 points each semester date (β = -3.09, p = .037). Though the regular school variable should represent the reference group, it was found that group differences were explained better when it was included in the model compared to when it was not. Attending a regular elementary school was found to be major disadvantage when it came to students' initial score and growth over time in this stage of modeling.

For students in magnet schools, initial status is on average 21.48 points lower than other students and the accelerated growth rate is nearly three points slower than other students (β = -2.79), but this slope does not reflect differences between student groups (p = .089). Attending a charter school is associated with a 4.07 point disadvantage when it comes to initial mathematics achievement and an accelerated growth rate of 1.66 points per semester, but neither value represents significant group differences (p = .830 and p = .606, respectively). They are not included in the next stage of modeling.

Once school type variables are added to race and CDI in the model, the variance is reduced from 2422.35 to 1859.32. This means that just over twenty-three percent (23.2%) of the variation in mathematics scores among students is explained by bringing students' school types into the model. However it also means that a small amount of error was introduced, as intercept variance increased from 1806.16 in the previous model. Variation among accelerated growth rates in mathematics scores is still significant and is explained mostly by race and regular school type. Bringing students' school types into the mathematics quadratic growth model also reduced the accelerated growth variance from 35.42 to 29.35 meaning that these predictors explain about eighteen and a half percent (18.6%) of the variance in the accelerated slope.

Table 24. Mathematics Results for the Conditional Model – Race, CDI, and School Type as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	489.74	9.75	50.25	1942	0.000
White	32.61	6.57	4.97	1942	0.000
Black	-12.34	5.99	-2.06	1942	0.039
Hispanic	-13.87	6.78	-2.05	1942	0.041
CDI	-4.39	1.11	-3.94	1942	0.000
Regular	-26.59	8.74	-3.04	1942	0.003
Magnet	-21.48	9.70	-2.21	1942	0.027
Charter	-4.07	18.11	-0.215	1942	0.830
Time					
Intercept	11.62	8.16	1.42	1942	0.154
White	-11.42	5.49	-2.08	1942	0.037
Black	8.55	5.00	1.71	1942	0.087
Hispanic	10.27	5.69	1.81	1942	0.071
CDI	2.21	0.94	2.35	1942	0.019
Regular	16.50	7.33	2.25	1942	0.024
Magnet	13.39	8.14	1.65	1942	0.100
Charter	-3.82	15.89	-0.24	1942	0.810
Time^2					
Intercept	2.08	1.64	1.26	1942	0.207
White	2.21	1.11	1.99	1942	0.046
Black	-2.00	1.01	-1.99	1942	0.046
Hispanic	-1.94	1.15	-1.69	1942	0.091
CDI	-0.53	0.19	-2.81	1942	0.006
Regular	-3.09	1.48	-2.09	1942	0.037
Magnet	-2.79	1.64	-1.70	1942	0.089
Charter	1.66	3.23	0.52	1942	0.606

Table 25. Variance Components for the Conditional Model – Race, CDI, and School Type as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	43.12	1859.32	1498	21.48.56	0.000
Time Slope	26.29	691.39	1498	1771.59	0.000
Time^2 Slope	5.42	29.35	1498	1809.19	0.000
Level-1Error	20.91	437.39			

Race, CDI, School Type, and CSR program as predictors. The next stage of modeling introduced the CSR programs as predictors. As is shown in Tables 26 and 27, the average initial mathematics score for the average student, regardless of gender is 468.54 and there is significant variation around this initial score, $\pi = 1867.32$, $p < .001$. Over time the accelerated growth rate for the average student is -1.54 points per semester, indicating that over time average student growth decreases about one and a half points per semester when the CSR programs are considered, and there is significant variation around the accelerated growth rate, $\pi = 28.84$, $p < .001$, though it does not reflect significant growth, $p = .101$.

Initially, the average mathematics score for White students is 41.83 points higher than the average non-White student. Over time the accelerated growth rate for the average White student's score increases by 3.47 points per semester. For initial mathematics scores, Black students are at a 3.06 point disadvantage, but this does not reflect significant group differences ($p = .453$). Over time, the accelerated growth rate decreases by an average of 1.04 points for Black students, but it is not significant ($p = .134$) and it is not retained as a predictor in the next model. CDI is still associated with a 4.36 point disadvantage when it comes to initial mathematics scores and a half point ($\beta = -0.44$) slower accelerated growth rate. While attending a regular school still predicts differential initial achievement ($\beta = -11.36$, $p = .011$), it no longer predicts differential accelerated growth ($\beta = -1.34$, $p = .077$) and is not included in the final model.

When it came to the CSR intervention programs, students from AC schools are at a 3.83 point disadvantage compared to students in other schools, and the quadratic growth rate for these students increases about one and a half points each semester date (β = 1.60, p = .052), while controlling for race, community disadvantage, and school type. AC school attendance did not predict differential initial status, but it approached significance for accelerated growth rate and is retained in the final model. For students in SFA schools, initial status is on average 9.74 points lower than other students and the accelerated growth rate for this student group is about half a point slower than other students (β = -0.65), but neither initial status nor slope reflect differences between student groups (p = .059 and p = .465, respectively) and this group is removed from the final model. Attending an ASP school is associated with non-significant 0.61 point disadvantage when it comes to initial mathematics achievement and a non-significant accelerated growth rate of 1.42 points per semester. Initially, the ASP group was removed from the model, but this introduced a lot of error and it was retained in the final model.

Once the CSR programs are added to the model, the variance is reduced from 2422.35 in the unconditional model to 1867.32. This means that just about twenty-three percent (22.9%) of the variation in mathematics scores among students is explained by bringing CSR programs into the model. As with the last model, a small amount of error was also introduced, as intercept variance increased from 1859.32 in the previous model. Variation among accelerated growth rates in mathematics scores is still significant. Bringing students' CSR intervention program types into the quadratic model also reduced the accelerated growth variance from 35.42 to 28.84, meaning that these predictors explain about eighteen and a half percent (18.6%) of the variance in the accelerated slope.

Table 26. Mathematics Results for the Conditional Model – Race, CDI, School Type, and CSR Program as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	468.54	5.54	84.44	1942	0.000
White	41.83	5.06	8.27	1942	0.000
Black	-3.06	4.07	-0.75	1942	0.453
CDI	-4.36	1.13	-3.87	1942	0.000
Regular	-11.36	4.47	-2.54	1942	0.011
AC	-3.83	4.84	-0.79	1942	0.430
SFA	-9.74	5.17	-1.88	1942	0.059
ASP	-0.61	5.01	-0.12	1942	0.904
Time					
Intercept	28.85	4.65	6.20	1942	0.000
White	-18.40	4.23	-4.33	1942	0.000
Black	3.28	3.43	0.96	1942	0.338
CDI	1.76	0.95	1.85	1942	0.064
Regular	7.63	3.74	2.04	1942	0.041
AC	-6.65	4.07	-1.64	1942	0.102
SFA	5.57	4.36	1.28	1942	0.202
ASP	-5.13	4.21	-1.22	1942	0.224
Time^2					
Intercept	-1.54	0.94	-1.64	1942	0.101
White	3.47	0.86	4.02	1942	0.000
Black	-1.04	0.69	-1.50	1942	0.134
CDI	-0.44	0.19	-2.27	1942	0.023
Regular	-1.34	0.76	-1.77	1942	0.077
AC	1.60	0.82	1.94	1942	0.052
SFA	-0.65	0.88	-0.73	1942	0.465
ASP	1.42	0.85	1.66	1942	0.096

Table 27. Variance Components for the Conditional Model – Race, CDI, School Type, and CSR Program as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	43.21	1867.32	1498	2150.46	0.000
Time Slope	26.03	677.63	1498	1766.59	0.000
Time^2 Slope	5.37	28.84	1498	1804.31	0.000
Level-1Error	20.92	437.79			

Final model. In the final model, after the non-significant variables were removed, the intercept represents the average mathematics achievement for students of color of any gender with average community disadvantage who did not attend an AC or ASP school. As is shown in Tables 30 and 31, the average initial mathematics score for the average student of color of any gender with average community disadvantage, not attending an AC or ASP school is 453.02 and there is significant variation around this initial score, π =1894.30, $p < .001$. Over time, the accelerated growth rate for the average student is -3.51 points per semester, indicating that over time average student accelerated growth slows about three and a half points per semester, and there is significant variation around this rate, $\pi = 29.13$, $p < .001$.

Initially, the average mathematics score for White students is 43.05 points higher than the average non-White student. Over time the accelerated growth rate of the average male student's reading score is 4.404 points faster than the average non-White student per semester. When it comes to students' CDI, it is negatively related to initial status (β = -4.49, $p < .001$) and negatively related to accelerated growth ($\beta = -0.45$, $p = .019$) when controlling for race and intervention program, indicating that higher community disadvantage is associated with lower initial mathematics achievement and slower growth over time. Results indicate that students in an AC school have initial mathematics scores which are half a point higher on average than other students, which is not a significant difference ($p = .896$), but being in an AC school is associated with an accelerated growth rate of 1.76 points per semester, which is significant. In addition, students in ASP schools have a 3.40 point advantage over other students when it comes to initial mathematics achievement, which does not reflect significant group differences (p = .430), but the accelerated growth rate for these students is 1.62 points per semester, which does reflect significant group differences.

In the final model the variance is reduced from 2422.35 to 1894.30. Approximately twenty-two percent (21.7%) of the variation in mathematics scores among students is accounted for by students' race, CDI, and attendance at an AC or ASP school. At this stage, variation among accelerated growth rates in mathematics scores is explained by race, CDI, and attendance at an AC or ASP school and is still significant. In the final model, accelerated growth variance was reduced from 35.42 to 29.13, meaning that the predictors explain almost eighteen percent (17.8%) of the variance in the accelerated slope. While the final quadratic growth model contains no non-significant predictors and error variance has been reduced from the unconditional model, a great deal of variance in student reading scores (more than 78%) is still left unexplained. As with reading, this suggests that there are other variables not included in this analysis that account for differential achievement and growth rate which might be explored in future research.

Table 28. Mathematics Results for the Final Conditional Model – Race, CDI, and CSR Program as Predictors

Level-2	*Coefficient*	*Std. Error*	*t-ratio*	*df*	*p-value*
Intercept					
Intercept	453.02	2.82	160.45	1945	0.000
White	43.05	4.45	9.67	1945	0.000
CDI	-4.49	1.13	-3.99	1945	0.000
AC	0.54	4.10	0.13	1945	0.896
ASP	3.40	4.31	0.79	1945	0.430
Time					
Intercept	39.49	2.34	16.67	1945	0.000
White	-20.00	3.74	-5.35	1945	0.000
CDI	1.85	0.95	1.95	1945	0.051
AC	-8.93	3.44	-2.60	1945	0.010
ASP	-7.31	3.62	-2.02	1945	0.043
Time^2					
Intercept	-3.51	0.48	-7.32	1945	0.000
White	4.04	0.76	5.33	1945	0.000
CDI	-0.45	0.12	-2.35	1945	0.019
AC	1.76	0.70	2.53	1945	0.012
ASP	1.62	0.73	2.21	1945	0.027

Table 29. Variance Components for the Final Conditional Model – Race, CDI, and CSR Program as Predictors

Random Effect	*Standard Deviation*	*Variance Component*	*df*	*Chi-square*	*p-value*
Initial Status	43.52	1894.30	1501	2163.90	0.000
Time Slope	26.20	686.20	1501	1770.83	0.000
Time^2 Slope	5.40	29.13	1501	1808.45	0.000
Level-1Error	20.93	438.12			

CHAPTER FIVE

DISCUSSION

This study was conducted with the goal of answering three key research questions: Are student outcomes on *TerraNova* mirrored by outcomes on *Supera* (the Spanish language *TerraNova* test)? How do the three CSR programs compare in terms of how they impact student growth over time based on *TerraNova* test scores? What are the effects of student- and school-level factors on the outcomes of these comprehensive school reform initiatives? In order to answer these questions, the researcher first conducted a series of *t*-tests, a repeated-measures analysis of variance, and then built a 2-level quadratic growth model that explored the factors that impact student achievement over time.

The preliminary analyses revealed a surprising and unique relationship between Spanish-speaking and English-speaking student performance on the two test versions. The hierarchical modeling showed that there is variation when it comes to student growth in reading and mathematics as measured by *TerraNova* standardized test scores. It was also shown that student- and school-level characteristics such as gender, race, motivation, and community socio-economic status do have an impact on achievement. Further, the results demonstrated that the three CSR programs, America's Choice, Success for All, and Accelerated Schools Project all have differential effects on student growth and performance in the two subject areas. The following sections will highlight the significant findings of the analyses, discuss the implications of these findings, describe the limitations of the present study, and lastly provide some recommendations for future research on comprehensive school reform.

Significant Findings

Supera and *TerraNova* Comparison.

The first research question asked whether or not there were differences between student achievement on the *Supera* test and the *Terra Nova* test. While it was expected that students taking *TerraNova* would outperform students taking *Supera* in both reading and mathematics and reading, the results revealed a more complex relationship between the student group and their performance.

T-tests. Reading score comparisons revealed that student scores only differ for the first test taken in the fall of first grade. Scores on *Supera* and *TerraNova* are not statistically different. Further, *TerraNova* average scores are above *Supera* average scores on the first and third tests, but the reverse is true for the second and fourth test dates. This suggests that students in the two test groups (English-speakers and Spanish-speakers) are learning reading content at different speeds and in different ways.

On the other hand, mathematics scores showed a completely different trend. Whereas for reading, students taking *Supera* and *TerraNova* were scoring similarly, the two groups were statistically different for mathematics. At all test dates, students taking *TerraNova* outscored students taking *Supera.* These findings are consistent with expectations, as well research on language equivalence (Jurado et al., 2012; van de Vijver &Tanzer, 1997). However, it is also worth noting that the fourth and final pair of scores, while significantly different, approached non-significance. This suggests that the gap between the groups' mathematics performance begins to close after a couple years of English instruction.

Repeated-measures ANOVA. The results of the *t*-tests raised new questions about the differences between the *Supera* test group and the *TerraNova* test group. A repeated-measures analysis of variance (ANOVA) was conducted for mathematics and reading scores that compared English- and Spanish-speakers average test performance over time. The results of the analysis revealed a significant effect of time on student achievement in both subjects. This was unsurprising due to the expectation that some amount of learning should have occurred between the first testing time point and last testing time point two academic years later.

What was more interesting was the finding that there is an interaction between the test taken and this growth over time for reading, but not for mathematics. While mathematics scores for both groups grow in a more linear fashion on parallel tracks, reading scores seem to change at different rates. Average *Supera* scores increase rapidly between the first and second test, more slowly between the second and third, and then faster again between the third and fourth. As the time between the second and third test is summer break for public school students, the slowed growth during this period is unsurprising and suggests that more could be done to encourage at-home summer reading activities for Spanish-speaking students, in particular.

Quadratic Growth Model.

The second and third research questions asked how do the three CSR programs compare in terms of how they impact student growth over time based on *TerraNova* test scores and what are the effects of student- and school-level factors on the outcomes of these comprehensive school reform initiatives? Overall, it was found that the rate of student growth in mathematics and reading does vary and that many factors, from gender to race to the CSR program in place have a significant impact on the rate of growth.

Average Growth in Reading and Mathematics.

In the first stage of hierarchical modeling, student scores serve as the outcome and only the growth variables (linear: Time and quadratic: Time^2) are included. The results of this Level-1 analysis describe the average growth of all students in the sample. Overall, students scored higher in reading than in mathematics, with average initial reading scores that were nearly 40 points higher than average initial mathematics scores. While immediate growth is faster for reading (40 points per semester) than for mathematics (30 points per semester), over time the accelerated growth rate for reading slows more than it does for mathematics. Reading growth rates slow approximately four and a half points over time and mathematics growth rates slow a little over a point and a half over time.

The unconditional models showed that there was significant variation when it came to student achievement growth in mathematics and reading. Additionally, the unconditional models contained a significant amount of variance in the outcomes. It was the goal of subsequent Level-2 models to explain some of the variation in student scores. Many student- and school- level factors were sequentially added to the model, but ultimately only a few key characteristics were found to impact student growth in reading or mathematics. For reading, the student characteristics of gender, race, and reading motivation were found to significantly impact student achievement and growth over time; only one CSR program, America's Choice was found to have a significant impact on the rate of growth. For mathematics, the student-level characteristic of race and the school-level characteristic of community disadvantage (a composite measure of SES) significantly impacted achievement and the rate of growth over time. Further, two CSR programs, America's Choice and the Accelerated Schools Project were found to have a significant relationship with student growth in mathematics.

Significant Predictors for Reading Achievement.

Gender. One student characteristic that was found to have a significant relationship to student academic performance and growth over time was gender. On average, boys score about 12 points lower than girls in reading. Over time, boy's rate of growth in this subject is almost two points slower than it is for girls. This difference suggests that reading programs may need to more to target boys, especially. These results are contrary to the original expectation that there would be no effect of gender on reading scores, but are in line with research on the differences between boys and girls when it comes to verbal skill acquisition (Logan & Johnston, 2010). Logan and Johnson (2010) explain that differential achievement may be attributed to differences in reading strategies between boys and girls, and that a more phonological (letter-sound oriented) approach may lead to greater engagement from boys.

Race. Students' race was also found to be a significant predictor of both initial reading achievement, as well as the rate of growth over time. On average, White students score almost 30 points higher than non-White students on the reading test. Over time, the rate of growth in this subject is about two points faster for White students than it is for students of color. These findings were consistent with the literature regarding the achievement gap that exists between minority and non-minority students (Gorey, 2009; Kyung, 2011; Reardon, Greenberg, Kalogrides, Shores, & Valentino, 2012; Slavin & Madden, 2001). It is evident that reform efforts insufficiently account for this particular demographic advantage and that more needs to be done to ensure that minority student populations are receiving the additional academic support services they need in order to narrow the chasm between their scores and those of non-minority students.

Motivation. The final student-level predictor that was found to be a significant predictor of student achievement and growth in reading was reading motivation. Motivation was measured using a student self-report survey that asked students to rate the truth of a series of statements regarding their interest in and ease of learning reading. Results of the quadratic model revealed that on average, that reading motivation has a positive relationship to initial achievement. However, over time, it has a negative relationship to the rate of growth. These findings are consistent with research on student motivation and reading achievement, but also suggest that there is more going on (Logan & Johnston, 2010). It may be the case that high motivation does not necessarily lead to high achievement and that interest in a subject does not automatically lead to academic success in that area. In fact, while confidence in one's abilities can bolster test performance, overconfidence can have the opposite effect and lead to careless mistakes.

America's Choice (AC). One of the central goals of this study was to identify which of three CSR programs has the greatest impact on student mathematics and reading achievement. When it came to reading scores, only one program was found to significantly explain differences in students' rate of growth over time. Students in an America's Choice school were found to grow at a rate nearly three points faster than other students. This is enough to compensate for gender or racial differences that predict a score deficit. While these findings are consistent with research demonstrating that AC places an emphasis on developing students' language arts (writing and reading) skills (Poglinco et al., 2003), it is worth noting that the AC group had significantly higher initial scores than other students (more than 17 points). This suggests that students in those schools may have started off with less need for reading remediation than those in other schools. It may also indicate that schools that already emphasize developing literacy skills may be drawn to initiatives such as America's Choice that have this as a core program goal.

Significant Predictors for Mathematics Achievement.

Race. The only student-level characteristic found to be a significant predictor of students' mathematics achievement and expected growth over time was race. On average, White students have an initial mathematics score which is more than 40 points higher than that of non-White students, reflecting a gap that is even greater than it was for reading scores. Over time, this gap only widens, as White students grow at a rate that is more than four points faster than that of students of color. As with reading, these findings support the existing literature regarding the minority achievement gap (Gorey, 2009; Kyung, 2011; Reardon, Greenberg, Kalogrides, Shores, & Valentino, 2012; Slavin & Madden, 2001). However, they also highlight the need of reform interventions to emphasize mathematics more than they are. Many programs place literacy and language arts skills as their focal point, but this may be hurting students when it comes developing mathematical reasoning skills.

Community Disadvantage Index (CDI). While CDI was not found to significantly predict differences between students' reading achievement, it was found to explain differential mathematics performance and growth over time. CDI is a composite measure of socio-economic status (SES) which factors in income, education level, employment, and single-parent status (Ball, Cohen, & Rowan, 2010). It ranged from 0 to 5, with 5 representing the greatest disadvantage and 0 representing the least disadvantage. CDI had a negative relationship to initial mathematics achievement, with greater disadvantage associated with a lower score on the first test. Over time, it maintains a significant negative relationship to the rate of student growth in mathematics. These findings reiterate that economically at-risk populations are at an academic disadvantage and provide further evidence that students from struggling communities need additional academic support to offset the negative impact of their socio-economic environment.

*America's Choice (AC) and (Accelerated Schools Project) ASP.*Ultimately, the goal of the analysis was to understand not only the different student characteristics impacting achievement and growth, but also to identify which CSR program(s) impact these trends, as well. When it came to mathematics achievement, two programs were found to significantly explain differences in students' rate of growth in this subject: America's Choice and the Accelerated Schools Project. While there were no significant differences between the students in these two groups and other students in initial mathematics achievement, both predicted differences in the accelerated rate of growth over time.

America's Choice students had a growth rate which was more than one and three quarter of a point faster than that of other students. Not only do these results support the articulate goals of the program (Poglinco et al., 2003), they also expand on earlier research which found less distinction between AC schools and other schools when it came to mathematics achievement (Supovitz et al., 2002). Accelerated Schools Project students had a growth rate which was more than a point and a half faster than that of other students. Despite the fact that ASP did not explain differential reading achievement growth, these findings are consistent with Byrd and Finnan's (2003) findings that ASP produces achievement gains in multiple subject areas. However, they also ran contrary to the expectation that there would be no difference between ASP and SFA schools when it came to mathematics achievement and growth.

Implications of the present study

The results of this study reveal that certain student populations may require a more targeted approach to academic intervention. When it comes to reading interventions, the findings suggest that boys, in particular, and minority students, in general require more academic support than their peers. Though this is purported to be an existing goal of many CSR programs, these groups continue to perpetually underperform. When it comes to mathematics, low socio-economic status and minority students are also at a disadvantage that does not seem to be adequately addressed by CSR interventions. The findings of this study should serve to emphasize the need for reforms that more thoughtfully address the potential detrimental academic effects associated with belonging to an at-risk population.

Limitations of the present study

There are several limitations to the scope of this study. The greatest limitation to this study was the over-specification of the growth model. This was a result of choosing to include and then remove reference groups from the model, which was done in order to show the specific impacts of different group membership. Future iterations of this work should reconsider how to dummy code variables such as race, as well as how these variables are introduced to the model.

Another concern is that despite using a diverse sample, the generizability of the results to other samples will be limited. Every school context is different and the outcomes of this analysis will only truly reflect the experiences of the participants. Despite efforts made to include a representative and balanced sample of schools and students, the influence of confounding factors is inescapable, especially in a school setting where it is nearly impossible to control all variables. Some confounding variables include the threat of item bias in the *Supera* test, differences in the level of implementation and variations in education leadership within each school, differences in the fidelity of implementation, variations in student, faculty, and parent “buy-in.” These factors should be kept in mind for the interpretation of the results this study.

In addition, the impact of historical context is inextricable from student outcomes. For example, it is important to note that data collection for this study ended almost ten years ago and the educational climate, as well as the content of the curriculum, may have changed enough during that time to alter the potential impact of one of these CSR programs. Further, this study is limited by the variables included in the SII. While *TerraNova* test data provides a useful measure of student achievement to conduct the analyses, grade data would have provided the opportunity to expand the meaning of academic growth to include more than one scholastic outcome.

Another way this study is limited is due to the role of the researcher as a secondary data analyst. It was a challenge to understand the dataset with nearly the same depth as the primary researchers. Access to the data files does not provide the intimate knowledge of the study and its many elements that Ball, Cohen, and Rowan (2010) were able to develop over the four year span of the SII. As always, missing data also threatens the validity of outcomes.

Recommendations for future research

Based on the unexpected and complex relationship between many student-level variables and reading and mathematics achievement, a number of questions remain that require further investigation. The *Supera*/*TerraNova* test comparison revealed a unique trend in reading growth that suggests more research should be done which explores the differences between English-speaking students and Spanish-speaking students when it comes to achievement over time and on a broader scale, as the present data set did not contain a large enough sample of Spanish-speakers to fully explore these differences. Further, the differences between these two groups bring up questions about language equivalent tests and what the true intent of the Spanish-language test is, both for students and for teachers. In addition, there is a question raised regarding how the intent of the test shapes its development.

The results of the quadratic modeling also unearthed several variables that impact student achievement and growth over time which warrant further investigation. Research on the role of different learning styles, student motivation, and SES would enrich researchers', educators', and policy makers' understanding of why certain programs are effective for some students, but not for others. In addition, it might be useful to include different subject content areas (such as writing, history, and science) in future research on comprehensive school reforms so that the full scope of interventions might be better understood.

REFERENCE LIST

Accelerated Schools Project (2002).About the Accelerated Schools Project. *TC.Columbia.edu*. Retrieved March 15, 2013 from http://www.tc.columbia.edu/accelerated/about.html.

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, *36*(5), 258-267.

Ball, D. L., Cohen, D. K., & Rowan, B. Study of Instructional Improvement (SII). ICPSR26282-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-20.

Blank, R. K. (2011).Closing the achievement gap for economically disadvantaged students? Analyzing change since No Child Left Behind using state assessments and the national assessment of educational progress. *Council of Chief State School Officers*.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2005). The national randomized field trial of Success for All: Second-year outcomes. *Success for All Foundation.*

Byrd, S. P., & Finnan, C. (2003). Expanding expectations for students through accelerated schools. *Journal of Staff Development*, 24(4), 48-52.

Burke, L. M., & Foundation, H. (2012).The student success act: Reforming federal accountability requirements under No Child Left Behind. Web Memo. No. 3461. *Heritage Foundation*.

Curriculum Reform (2013).America’s Choice.*Curriculumreform.wikispaces.com*. Retrieved March 15, 2013 from http://curriculumreform.wikispaces.com.

Evans-Andris, M., & Usui, W. M. (2008). Comprehensive school reform & student Achievement in Kentucky middle schools. *Middle Grades Research Journal*, 3(2), 51-69.

Gorey, K. M. (2009). Comprehensive school reform: Meta-analytic evidence of Black-White achievement gap narrowing. *Education Policy Analysis Archives*, *17*(25), 1-14.

Hurley, E. A., Chamberlain, A., Slavin, R. E., & Madden, N. A. (2001). Effects of Success for All on TAAS reading: A Texas statewide evaluation. Report no. 51. *Center for Research on the Education of Students Placed at Risk.*

Jurado, M., Cumba-Aviles, E., Collazo, L. C., & Matos, M. (2006). Reliability and validity of a Spanish version of the Social Skills Rating System-teacher form. J*ournal of Psychoeducational Assessment*, *24*(3), 195-209.

Kyung Eun, J. (2011). Thinking inside the box: Interrogating No Child Left Behind and Race to the Top. *KEDI Journal of Educational Policy*, *8*(1), 99-121.

Logan, S., & Johnston, R. (2010).Investigating gender differences in reading. *Educational Review*, *62*(2), 175-187.

Mac Iver, M. J. (2009). Urban middle-grade student mathematics achievement growth under comprehensive school reform. *Journal of Educational Research*, *102*(3), 223-236.

Mumme, J. & Weissglass, J. (1990). Improving mathematics education through school-based change. In N.A. Fisher, H.B. Keynes, & P.D. Wagreich (Eds.), *Mathematicians and Education Reform, 1989-1990* (pp. 3-23). American Mathematical Society.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*(2), 365-397.

Poglinco, S. M., Bach, A. J., Hovde, K., Rosenblum, S., Saunders, M., & Supovitz, J. A. (2003). The heart of the matter: The coaching model in America's Choice schools. *Consortium for Policy Research in Education*.

Raudenbush, S.W. & Bryk, A.S.(2002). *Hierarchical Linear Models*. Thousand Oaks, CA:Sage Publications.

Reardon, S. F., Greenberg, E., Kalogrides, D., Shores, K. A., & Valentino, R. A. (2012). Trends in academic achievement gaps in the era of No Child Left Behind. *Society for Research on Educational Effectiveness.*

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal*, *44*(2), 252-297.

Slavin, R. E., & Madden, N. A. (2001). Reducing the gap: Success for All and the Achievement of African-American and Latino students. *Center for Research on The Education of Students Placed at Risk.*

Slavin, R. E., & Madden, N. A. (2004). Success for All/Roots & Wings: Summary of research on achievement outcomes. Report no. 41 (revised). *Center for Research on the Education of Students Placed at Risk.*

Success for All Foundation (2012).About us. SuccessforAll.org. Retrieved March 15, 2013, from www.successforall.org/About-Us/.

Supovitz, J. A., Taylor, B., & May, H. (2002). The impact of America's Choice on student performance in Duval County, Florida. *Consortium for Policy Research in Education*.

Tivman, T., & Hemphill, L. (2005).Comparing four literacy reform models in high-poverty schools: Patterns of first-grade achievement. *Elementary School Journal*, *105*(5), 419- 441.

Tushnet, N. C., Flaherty, J., & Smith, A. (2004). Longitudinal assessment of comprehensive school reform program implementation and outcomes: First-year report. *US Department of Education.*

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Europeenne de Psychologie Appliquee,* 54, 119–135.

What Works Clearinghouse (2012). Success for All: What Works Clearinghouse intervention report. *What Works Clearinghouse*.

Zimmer, R., Gill, B., Razquin, Booker, K., & Lockwood, J.R.. (2007). State and local implementation of the "No Child Left Behind Act." Volume I--Title I: School choice, supplemental educational services, and student achievement. A Report from the National Longitudinal Study of "No Child Left Behind" (NLS-"NCLB"). *US Department of Education.*

VITA

Gabriella S. Gorsky was born in Chicago, IL. She graduated from DePaul University Chicago in 2011 with a BA in Psychology. In 2011, she began working towards her Masters in Research Methodology from Loyola University Chicago. Currently, Gabriella works as a research assistant in the Office of Research, Evaluation, and Assessment at Evanston Township High School in Evanston, IL. In the fall of 2013 she will begin doctoral work in the Measurement, Statistics, and Research Design program at the University of Washington, Seattle.